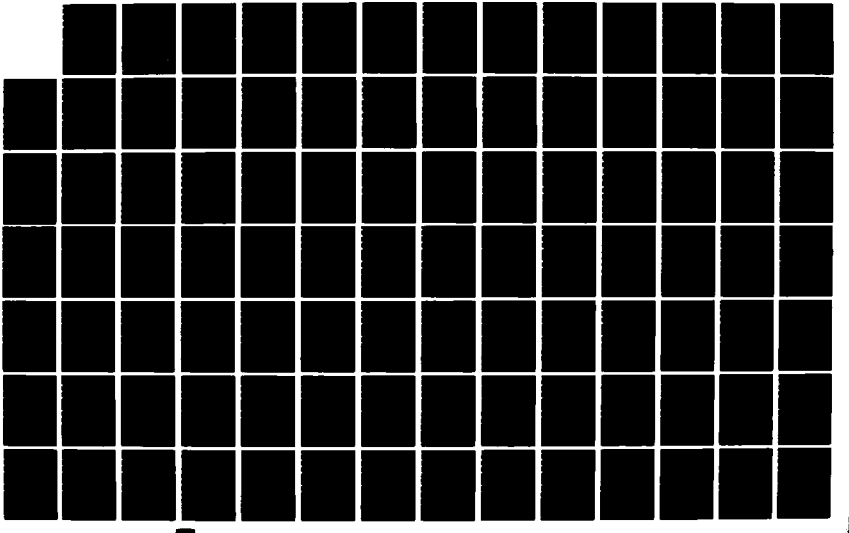
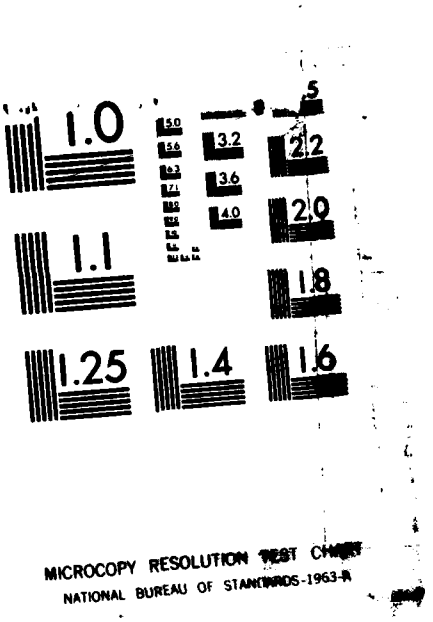
AD-A166 374 A COMPUTER SIMULATION OF THE L-ARABINOSE GENE-ENZYME COMPLEX WITH AN ANAL (U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH B L GEORGE 1985 1/2

UNCLASSIFIED AFIT/CI/NR-86-6D F/G 6/1 NL




1.0
1.1
1.25
1.4
1.6
2.2
2.0
1.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

AD-A166 374

1. REPORT NUMBER: AFIT/CI/NR 86-6D
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*: A Computer Simulation of the L-arabinose Gene-Enzyme Complex an Analysis of Its Control Methodology
5. TYPE OF REPORT & PERIOD COVERED: THESIS/DISSERTATION
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Bruce L. George
8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT: The Ohio State University
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR, WPAFB OH 45433-6583
12. REPORT DATE: 1985
13. NUMBER OF PAGES: 145

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*
15. SECURITY CLASS. *(of this report)*: UNCLASS
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

LYNN E. WOLAVER 4 April 86
Dean for Research and Professional Development
AFIT/NR, WPAFB OH 45433-6583

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

DTIC ELECTE S APR 1 0 1986 D E

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

086 4 8

+6C
6

A Computer Simulation of the L-arabinose Gene-Enzyme Complex with an Analysis of Its Control Methodology

By

Bruce L. George, Ph.D.

The Ohio State University, 1985

Professor Richard M. Campbell, Adviser

This research investigated mathematically aspects of the "Demand Theory of Gene Regulation," which relates the evolution of control of gene activity to the environmental pressure upon the organism. The specific goal was to utilize an engineering systems approach to quantify some portions of this theory and examine the energy cost to the organism of alternative strategies of genetic control.

A systems theory approach was taken to represent the biological system by a linear time invariant realization. The L-arabinose gene-enzyme complex of E. coli was simulated on the computer using an eight state space model. This operon is regulated by both a repressor and an activator thus combining both negative and positive control. The first six states of the model represented protein/DNA interactions, while the final two states represented the concentrations of the repressor and activator proteins. All inputs to the system (RNAP, cAMP and L-arabinose) were considered as step

Cont'd

86 4 8 084

Cont'd

inputs. The output (L-arabinose isomerase specific activity) was related directly to the activity of the DNA in the biological system. The stability, controllability, frequency response and state space relationships of the system model were studied. Coefficients in the system's equations were optimized based upon least square error criteria using as expected values the uninduced and induced values of L-arabinose isomerase specific activity as reported in the literature. Four models were developed and then compared to published biological data. Residual analysis on a set of biological variables was used to differentiate among the models and select the model of best fit.

A model utilizing repressor only control was also developed for the L-arabinose system. The resulting five state model was optimized and tested against reported data utilizing residual analysis.

A quadratic cost function was used to compare alternative control methodologies. Cost comparisons were made between the five state and eight state models. The energy cost to the cell was assumed to be proportional to the protein/DNA associations and the concentrations of the repressor and activator proteins. Results of the cost analysis show that a dual control system is more cost effective under most environmental conditions for this specific gene-enzyme complex.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

QUALITY INSPECTED 3

6

A COMPUTER MODEL OF THE L-ARABINOSE GENE-ENZYME COMPLEX OF *E. COLI* WITH AN ANALYSIS OF ITS CONTROL METHODOLOGY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the Graduate

School of the Ohio State University

By

Bruce Lee George, B.S.,M.S.,M.S.E.E.

* * * * *

The Ohio State University

1985

Dissertation Committee:

Richard M. Campbell

Paul A. Fuerst

Manjula B. Waldron

Herman R. Weed

Approved by

Richard M. Campbell

Adviser

Department of Electrical Engineering

ACKNOWLEDGMENTS

I sincerely thank Dr. Richard M. Campbell and Dr. Paul A. Fuerst for their guidance throughout this research. Thanks go to the other members of my advisory committee, Dr. Manjula B. Waldron and Herman R. Weed, for their suggestions and comments. To my fellow graduate students, I offer thanks for their technical assistance. To my wife, Joan, I offer special thanks for her support and understanding.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

VITA iii

LIST OF TABLES. vi

LIST OF FIGURES viii

CHAPTER . . . . . PAGE

1. INTRODUCTION. 1
   1.1 Overview 1
   1.2 Modeling Physiological Systems 4
   1.3 The L-Arabinose Gene-Enzyme Complex 6
   1.4 Alternative Models for Control of Genetic Systems 9

2. MODEL DEVELOPMENT 13
   2.1 Introduction 13
   2.2 Parameter Estimation. 21
   2.3 Eigenvalue Model Development:GENE1/2. 24
   2.4 Non Limited States Model Development: GENE1C/2C 27
   2.5 Limited Model Development: GENE3A/3B;GENEOA/OB 28

3. MODEL DISCRIMINATION 57
   3.1 Introduction 57
   3.2 Evaluation of the Models 61

4. SYSTEM ANALYSIS 74
   4.1 Introduction 74
   4.2 Eigenvalue Analysis 74
   4.3 Controllability 76
   4.4 Frequency Analysis 77
   4.5 State Space Analysis 79
   4.6 GENEOA/OB Analysis 80

5. REPRESSOR MODEL DEVELOPMENT 106

6. COST ANALYSIS 122

6.1 Introduction 122
6.2 Cost Analysis for the L-arabinose System . . 123

7. DISCUSSION 136
7.1 Conclusions 136
7.2 Discussion of the Models 138
7.3 Future Considerations 139

LIST OF REFERENCES 140

GLOSSARY 141

LIST OF TABLES

TABLE	PAGE
2.1 GENE1/2 Eigenvalue Sensitivity	40
2.2 GENE1/2 Eigenvalue Optimization (EV3) . . .	41
2.3 GENE1/2 Eigenvalue Optimization	42
2.4 GENE1/2 Eigenvalue Model	43
2.5 GENE1/2 B Parameter Readjustment	44
2.6 GENE1C/2C Parameter Sensitivity	45
2.7 GENE1C/2C Parameter Optimization	48
2.8 GENE1C/2C B Matrix Optimization	51
2.9 GENE3A/3B B Matrix Parameter Sensitivity . .	52
2.10 GENE3A/3B Parameter Sensitivity	53
2.11 GENE3A/3B Parameter Optimization	55
2.12 GENEOA/OB Coefficient Classes with Optimized Values	56
3.1 GENEOA/OB	65
3.2 GENE1/2	66
3.3 GENE1C/2C	67
3.4 GENE3A/3B	68
3.5 GENEOA/OB Residual Analysis	69
3.6 GENE1/2 Residual Analysis	70
3.7 GENE1C/2C Residual Analysis	71

3.8 GENE3A/3B Residual Analysis 72

3.9 Model Discrimination Residual Analysis . . . 73

4.1 Eigenvalues 101

4.2 GENE7A/7B Analysis 102

4.3 GENE7A/7B 103

4.4 GENE0A/0B and GENE7A/7B Residual Analysis . 104

4.5 GENE0A/0B Catabolite Repression Analysis . . 105

5.1 GENE4A/4B B Matrix Parameters 116

5.2 GENE4A/4B A Matrix Sensitivity Analysis . . 117

5.3 GENE4A/4B 118

5.4 GENE4A/4B Residual Analysis 119

5.5 GENE8A/8B 120

5.6 GENE8A/8B Residual Analysis 121

6.1 Cost Analysis 132

6.2 GENE9A/9B 133

6.3 GENE9A/9B Residual Analysis 134

6.4 Cost Analysis 135

LIST OF FIGURES

FIGURES	PAGE
1.1 L-Arabinose Gene-Enzyme Complex	7
2.1 Block Diagram of L-arabinose System	16
2.2 Original System Model	31
2.3 Simulation Program	32
2.4 Uninduced Response/Original Model	35
2.5 Induced Response/Original Model	36
2.6 Starting System Model	37
2.7 Real Jordan Normal Form of Starting Model . .	38
2.8 New A Matrix for Eigenvalue Model	39
3.1 GENEOA/OB	64
4.1 Controllability Matrix of Original Model. . .	83
4.2 Transfer Matrix, H(s), for GENEOA/OB	84
4.3 PBAD vs Pc for GENEOA	85
4.4 PBAD vs Pc for GENEOB	86
4.5 Free Protein vs Time for GENEOA	87
4.6 Bound Protein vs Time for GENEOA	88
4.7 Free Protein vs Time for GENEOB	89
4.8 Bound Protein vs Time for GENEOB	90
4.9 Uninduced Response of GENEOA	91
4.10 Induced Response of GENEOB	92

4.11 P1 vs P2 for GENEOA 93

4.12 P1 vs P2 for GENEOB 94

4.13 Output vs Free Protein for GENEOA 95

4.14 Output vs Free Protein for GENEOB. 96

4.15 Output vs araI for GENEOA 97

4.16 Output vs araI for GENEOB 98

4.17 Real Jordan Normal Form Of GENEOA/OB 99

4.18 New A Matrix for GENE7A/7B 100

5.1 Uninduced Response of GENE4A 111

5.2 Induced Response of GENE4B 112

5.3 GENE8A/8B 113

5.4 Uninduced Response of GENE8A 114

5.5 Induced Response of GENE8B 115

6.1 GENE9A/9B 128

6.2 Uninduced Response of GENE9A 129

6.3 Induced Response of GENE9B 130

6.4 Energy Cost vs Percent Induction
GENEOA/OB (Eight State Model) &
GENE9A/9B (Five State Model) 131

Chapter 1

Introduction

1.1 Overview

The purpose of this research was to investigate mathematically aspects of the "Demand Theory of Gene Regulation." According to this theory, control of gene activity evolves in response to the environmental pressure exerted upon the organism [1,2]. When the environment demands frequent expression of a particular gene, selection favors a positive form of gene control (eg. activator control). When a gene is rarely expressed in a particular environment, selection favors negative control (eg. repressor control). The specific goal of the present study was to utilize an engineering approach to quantify further some portions of this theory by examining the energy cost to the cell of alternative strategies of genetic controls. This was done by forming state space system representations for the alternative strategies and relating the cellular energy cost to the states of the representations. A form of the quadratic cost function from optimal control theory was used to evaluate the energy cost to the cell [3].

The L-arabinose system of E. coli was selected for study because it presents an interesting and comparatively well documented case. This system, like the tryptophan operon of E. coli, is under dual

control [1]. Dual control systems are relatively rare. Most operons are under only singular control such as the negatively controlled galactose operon or the positively controlled maltose operon [2]. The L-arabinose operon is regulated by both a repressor and an activator thus combining both negative and positive control [4]. The theory hypothesizes that this dual control came about because E. coli commonly inhabits two different environments, one of which has arabinose present as a carbon source and one that does not [1].

The L-arabinose gene-enzyme complex was simulated on the computer using an eight state space linear model. The first six states represented protein/DNA interactions and the final two states represented the concentrations of the repressor and activator proteins. The computer model was based on the current repressor/activator model for the A,B,C and D genes of the L-arabinose complex [4,5,6]. Where possible, data from the literature was used to develop the interrelationships between the states, and their relationships to the inputs and the outputs. Where data did not exist in the literature, the best engineering/biological estimates were made. All inputs to the system (RNAP, cAMP and L-arabinose) were considered as step inputs (representing a constant level of substance available to the cell) and the output (L-arabinose isomerase specific activity) was taken as a linear multiple of the PBAD/RNAP associations. Coefficients in the system's equations were optimized based upon a least square error criterion using as expected values the uninduced and induced values of L-arabinose isomerase specific

activity as reported in the literature [7]. Four models in all were developed and then compared to the published data. A nonlinearity (limiting the states of the system to non negative values) was introduced for physiological matching in two of the models. Residual analysis was used to differentiate among the models and select the model of best fit.

In order to compare alternative control methodologies, a hypothetical model utilizing repressor only control was developed for the L-arabinose system. The resulting five state model was tested against reported data utilizing residual analysis, and also compared to the dual control eight state model.

A cost analysis was performed on both models and comparisons made between the models. The cost analysis was based upon a cellular cost function that was designed to represent an energy cost to the cell. The energy cost to the cell was assumed to be proportional to the protein/DNA associations and the concentrations of the repressor and activator proteins. The cost function was evaluated under several conditions including various mutational circumstances. The repressor only control model was found to be almost twice as energy expensive as the dual control model. Thus, the results of the cost analysis suggest that a dual control system is more cost effective for the regulation of this specific gene-enzyme complex.

1.2 Modeling Physiological Systems

To accomplish this research, a mathematical model of the Gene-Enzyme complex had to be developed. In general, a mathematical model can be defined as a set of equations that approximately describe the behavior of a particular system [8]. The model should closely resemble the actual system and predict its behavior under varying conditions in so far as possible. The model is usually simplified, describing only some of the parameters of the real system. How well a model works and its correspondence to the system being modeled depends greatly on the assumptions made during the design of the model. The assumptions made depend primarily on the viewpoint of the modeler. The modeler decides what information is to be included in the model and how this information is going to be used. Although there are no rules for model building, the following represent some of the general steps to model building [8]:

1. Understand the natural history of the system to be modeled. Identify the dependent and independent variables of the system, the interrelationships between these variables and the inputs and outputs of the system. This was done for the L-arabinose system using, where possible, information available from the biological literature. A discussion of the variables chosen and their interrelationships is found in section 2.1.

2. Review the literature on systems modeling. Examine models and approaches used on systems similar to the one to be modeled. Examine the similarities of other systems' models and problems

encountered during their modeling. A review of the literature relevant to this study is found in section 1.4.

3. Determine the mathematical form of the model. Choose between a linear or a nonlinear set of equations. A linear model was chosen in this study to represent the L-arabinose system. The reasons for this selection are discussed in section 2.1.

4. Estimate the parameters of the model. Choose a method of parameter optimization. Design the model against a specific criterion. The parameter optimization for the model of the L-arabinose system is found in section 2.2.

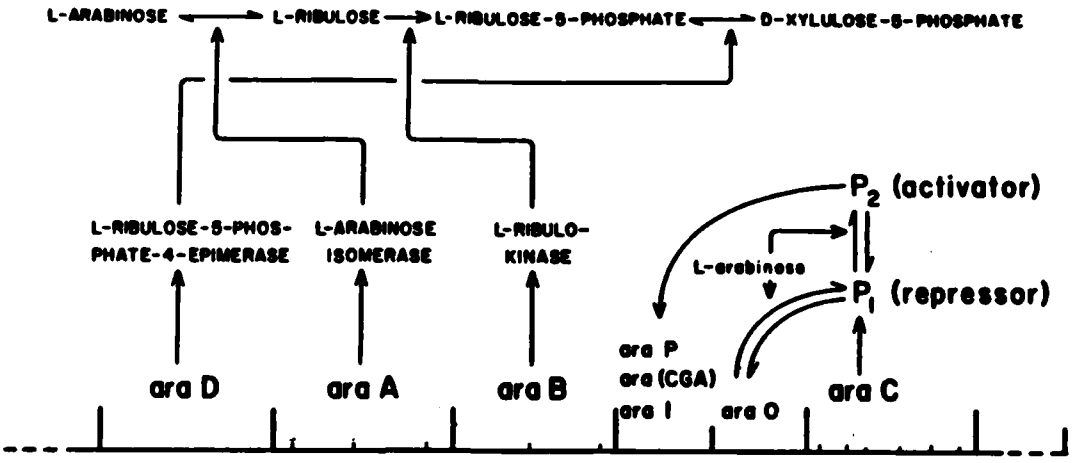
5. Test the model. Validate the model by examining the consistency of the model's output under known experimental conditions other than those used to develop the model. Testing of the model for the L-arabinose system is found in Chapter 3.

Following such an approach assists the modeler in preparing a mathematical description of a system that both resembles the system's structure and mimics its function.

Lastly, in the words of Economos [9], "For a model of a system to be valuable in physiological research, it should be able to generate more information than the modeller put into it." A model should be able to duplicate a system's response to conditions which were not taken into account during development of the model, and it should provide new hypotheses on some unknown properties of the system that could be later verified by experimentation.

1.3 The L-Arabinose Gene-Enzyme Complex

The L-arabinose gene-enzyme complex has been extensively reviewed by several researchers [4,5,6]. The complex consists of at least five structural genes and one controlling gene. It has been studied most thoroughly in the bacterium Escherichia coli. The structural genes of the complex contain information for the enzymes required for the metabolism of the sugar, L-arabinose. L-arabinose is a carbon source of unpredictable availability to the bacterium. Genetic mapping has revealed that three of the structural genes (B,A,D) and the controlling gene (C) lie on one contiguous segment of the E. coli chromosome (Figure 1.1) [10]. Genes B, A, and D code for the enzymes L-ribulokinase, L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase respectively. Expression of genes araA, araB and araD is controlled through a set of genetic sites which are located on the chromosome between araC and araB. These controlling sites include the operator site, araO, the promoter site for the araC gene, Pc, and the promoter site for the three structural genes, PBAD [11]. Genes B,A, and D are transcribed as a single unit which means that when the promoter, PBAD, is activated, all three genes turn on production of messenger RNA. This messenger RNA (mRNA) is then used in the cell for the production of the enzymes [12].



L-arabinose Gene-Enzyme Complex
From (4)
Figure 1.1

The remaining two genes of the complex, araE and araF, are located elsewhere on the chromosome. These genes are involved in the production of the enzymes responsible for the transport of L-arabinose into the cell. Even though they are located at some distance from the rest of the complex, these two genes are also under the control of the regulatory gene, araC [4].

This gene-enzyme complex is regulated by both positive and negative control. Such dual control is rarely found in genetic systems. The classic Jacob-Monod Operon model for gene control uses only negative control in the form of a repressor substance to keep a gene turned off until its product is needed [13]. More recent research has identified some genes that are normally turned off until an activator substance appears to turn on the transcription of the gene [14]. Genes that require an external activator to become transcriptionally active are said to be under positive control. In the

L-arabinose gene-enzyme complex, the controlling gene, araC, codes for a protein P1 that acts as a repressor for activity at both Pc and PBAD. P1 binds at the operator site, araO, and has the effect of turning off both Pc and PBAD. The inhibition of Pc is easily explained biochemically, since the binding site of P1 to araO overlaps the binding site of RNA Polymerase, RNAP, to Pc. RNAP is the enzyme which is used to produce messenger RNA from the coding DNA which makes up the gene complex. Thus, the repression of araC is due to competition for binding on the DNA between P1 and RNAP. If P1 is bound at araO, it prevents the attachment of RNAP with the consequent halt in transcription of mRNA from the araC gene [15].

The repression of the activity of the B, A, and D genes is not as easy to explain. PBAD and araO are far removed physically from each other on the chromosome. The concept of competitive binding between the repressor and RNAP probably does not apply. To date, there has been no satisfactory explanation for the known repressive effect of P1 on PBAD [16].

The repressor protein P1 is in dynamic equilibrium with another protein P2. P2 acts as an activator for PBAD. The effect of L-arabinose as an inducer for the complex is twofold. First it tends to free P1 from araO increasing binding of RNAP at the promoter sites. Secondly, it shifts the equilibrium between P1 and P2 towards a higher concentration of P2 [17]. Removal of the repressor from araO is not enough to turn the B, A, and D genes on. The activator, P2, must bind to the DNA in the region of PBAD (araI) before full transcription, and

consequent enzyme production occurs [7].

In addition to P2, another activator molecule is required for full stimulation of the complex. Cyclic adenosine monophosphate (cAMP) with its receptor protein (CRP) binds to the DNA near araI and appears to aid the binding of P2 [18]. Recently, it has been proposed that cAMP-CRP and P2 function primarily together to aid in the binding of RNAP to PBAD [19].

1.4 Alternative Models for Control of Genetic Systems

In the past, several models have been developed to describe the control of genetic systems. Previous modeling efforts have been based on the use of nonlinear chemical kinetic equations [20,21,22,23]. These equations have usually been linearized by some approximation and then solved under the assumption of a steady state solution. For example, Savageau has developed an approach based upon a power law formulation of the kinetic equations connecting protein production and resource utilization [20]. This power law formulation linearizes the kinetic equations describing specific parameters through the use of logarithms and a change of variables.

Alberghina's model for ribosome and protein synthesis, chromosome replication and cell division in E. coli involved a set of differential equations that were intermittently applied during the growth phase of the bacteria [23].

A recent and extensive example is Tapaswi's model for transcription of genes and translation of mRNA to proteins during embryogenesis [21]. Utilizing eight nonlinear synthesis reactions, he modeled the production of the various forms of RNA and proteins necessary for genetic activity during embryogenesis. Again, solution of the equations involved a change of variables and linearization around a steady state solution.

A set of linear equations were used in this study as opposed to the chemical kinetic equations mentioned above. One reason for the use of linear equations is that more standard analysis techniques exist for linear systems,such as stability analysis, controllability analysis and frequency analysis. [24]. Also, the results presented by Haggerty [25] indicate that when the araC protein is the limiting factor for the L-arabinose system, expression of PBAD is directly proportional to the level of araC protein. This supports the selection of a linear model for the system as a linear approach is suggested when the experimental evidence suggests linearity [9].

A new approach to modeling genetic systems is represented by the Threshold model of Tchuraev [21]. This model uses a unique set of equations that represent concentrations of substances, activity of genetic material and the interactive effects of regulatory substances on the genetic material. The input variables of the model represent either the presence or absence of a threshold dose of various substances, thereby discretizing the genetic system. Using this formulation, this model produced oscillatory system dynamics for the

operation of an arbitrary operon. A discrete model, such as this one, was not chosen to represent the L-arabinose system because the biological literature reports a constant steady state output of the system and this discrete model shows a steady state oscillatory behavior.

Discretizing of the genetic system has also been attempted by other researchers. Neumann and Kreischer [26] and Bellmann et.al. [27] have developed discrete models for the transcription of the DNA in simple and more complex cells respectively. These models were not used because the inputs to the biological system are continuous and not discrete.

The research described here presents a different approach to the modeling of a genetic system. Unlike previous models, the model developed here attempts to describe a genetic system's behavior by considering the interactions of a segment of DNA with the various known regulatory substances, and then extrapolating the activity of the DNA to the production of the desired protein. In this model, the activity of different sites of the DNA are represented by states of the system. A linear set of first order differential equations were written to discribe the behavior of the system for this model. Unlike previous models, this model allows the energy cost of genetic activity to be directly related to the states of the system, and permits comparisons of the cost to be easily made between different systems. In addition, the use of this state space approach allows the development of similar models of the same system using alternative

methods of control.

Chapter 2

Model Development

2.1 Introduction

The basic simplifying premise followed in the model design was that an inherently nonlinear biological system could be reasonably approximated by a linear time invariant (LTI) representation. This supposition was supported by Haggerty's finding of a linear relationship between the concentration of the araC protein and the activity of PBAD [25]. The intent of using an LTI model was to facilitate system analysis and model manipulation. Silverman [28] has reviewed the techniques that can be used to develop a state space realization of a dynamic system. These methods, however, require an initial description of the input/output relationship, in the form of a system transfer function, for the formulation of the state space realization. A system transfer function has not been formulated for the L-arabinose system. This necessitates the formation of a state space representation from the descriptive data found in the biological literature.

The basic state space structure for a linear time invariant system can be written as [24].

$$\dot{X} = AX + BU \qquad (2.1)$$

$$Y = CX + DU \qquad (2.2)$$

where X = n x 1 vector of the states of the system

$\dot{X}$ = n x 1 vector of the first derivatives of the states of the system (dx / dt)

A = n x n matrix of constants describing the relationships among the states of the system

B = n x p matrix of constants describing the relationships between the states of the system and the inputs

U = p x 1 vector of the inputs to the system

Y = q x 1 vector of the outputs of the system

C = q x n matrix of constants describing the relationships of the outputs of the system to the states of the system

D = q x p matrix of constants describing the relationships of the outputs of the system to the inputs.

For the L-arabinose system, the following values are proposed; n equal to eight, p equal to three and q equal to one. It is also proposed that D is a zero matrix for this system as is discussed in the development of the equations.

For the L-arabinose complex, the following state variables were chosen.

x_1 = RNAP/PBAD associations

x_2 = RNAP/Pc associations

x_3 = P1/araO associations

x_4 = P2/araO associations

x_5 = P2/araI associations

x_6 = cAMP/CRP associations

x_7 = concentration of P1

x_8 = concentration of P2

The first six states represent the activities of the known regulatory sites for the L-arabinose operon [4,5]. The last two states represent the concentrations of the known regulatory proteins for the L-arabinose operon [4,5].

The activity of araO was separated into two state variables because the biological literature suggests the possibility of different effects for repressor and activator binding to this site [4,15,17]. The early research (reviewed in [4]) suggested that araO was the site of repression for PBAD and Pc. Lee [15], using DNA binding studies, found no direct evidence for araO repression of PBAD, but did find evidence that the araO site was the site of repression of Pc. Hahn and Schleif [17] proposed that P1 bound to araO represses both PBAD and Pc, but that P2 bound to araO represses only Pc. In addition, the presence of these two distinct states allowed for clearer interrelationships between states to be expressed. It has also

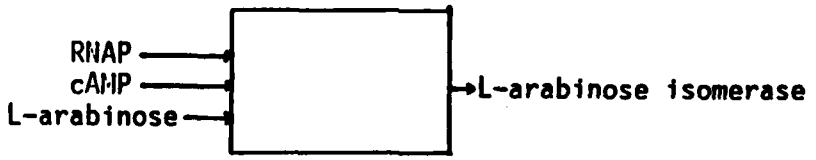
been recently reported that araO may actually be two distinct sites, one for the binding of P1 and one for the binding of P2 [19]. The state x_6 represents the associations of the cAMP/CRP complex with the DNA. The concentration of the free CRP and the interaction between the free cAMP and the CRP to form the complex were not modeled. It was assumed that the concentration of the free CRP and the rate of formation of the cAMP/CRP complex were not limiting factors of the system.

The following were chosen as inputs to the system (Figure 2.1).

u_1 = concentration of L-arabinose

u_2 = concentration of RNAP

u_3 = concentration of cAMP



Block Diagram of L-arabinose System

Figure 2.1

For simplicity, the three inputs to the system were considered as step inputs. This means that a constant supply of each input was available to the cell. The inputs were normalized to 1.0 so that any variation in the inputs could be handled by alterations to numbers in the input matrix (B matrix).

L-arabinose (u_1) was input to the induced cell only. The lack of L-arabinose input to the uninduced cell is represented by zeros in the first column of the B matrix. The differences between the B matrices represent the only differences in modeling between the uninduced and induced systems. The transport of L-arabinose into the cell as controlled by the araE and araF genes was not modeled in this study.

RNAP was input to the system to provide for Pc activity. Since RNAP activity is not specific for the L-arabinose system, it is expected the RNAP will always be available in the cell. The araC gene is autoregulated by its own product [29]. This autoregulation is believed to be simply the result of competitive binding between P1 at araO and RNAP at Pc [15].

The third input to the system, u_3, represented either an internally produced or externally provided source of cAMP.

The specific activity of L-arabinose isomerase was chosen as the output of the system (Figure 2.1). Of the three enzymes coded for by the araBAD operon, L-arabinose isomerase is the enzyme most commonly measured during experimentation. The units of specific activity are micromoles of ribulose formed per hour per milligram of protein. Differences in specific activity reflect differences in the amount of L-arabinose isomerase made by the cell. Thus, for this model, the specific acitvity of isomerase was assumed to be directly proportional to the amount of enzyme present and a simple linear relationship was assumed to exist between the activity of PBAD and the amount of

isomerase produced. The output of the system was taken as a simple multiple of the value of x_1 (RNAP/PBAD activity).

It is proposed that the equations representing the rates of change of the system's states can be written as:

$$\dot{x}_1 = -a_1 x_1 + a_2 x_5 \quad (2.3)$$

$$\dot{x}_2 = -a_3 x_2 - a_4 x_3 - a_5 x_4 + a_6 x_6 + b_1 u_2 \quad (2.4)$$

$$\dot{x}_3 = -a_7 x_3 - a_8 x_4 + a_9 x^7 \quad (2.5)$$

$$\dot{x}_4 = -a_{10} x_3 - a_{11} x_4 + a_{12} x_8 \quad (2.6)$$

$$\dot{x}_5 = -a_{13} x_3 - a_{14} x_5 + a_{15} x_6 + a_{16} x_8 \quad (2.7)$$

$$\dot{x}_6 = -a_{17} x_1 - a_{18} x_6 + b_2 u_3 \quad (2.8)$$

$$\dot{x}_7 = a_{19} x_2 - a_{20} x_3 - a_{21} x_7 + a_{22} x_8 - b_3 u_1 \quad (2.9)$$

$$\dot{x}_8 = -a_{23} x_4 - a_{24} x_5 + a_{25} x_7 - a_{26} x_8 + b_4 u_1 \quad (2.10)$$

With the output of the system written as:

$$y = c_1 x_1 \quad (2.11)$$

Equations 2.3-2.11 represent the dynamical equations for the L-arabinose gene - enzyme complex for an exponentially growing colony of E. coli cells. In equation (2.3), the rate of binding of RNAP to PBAD is assumed to be directly proportional to the amount of activator protein bound to araI, as the data suggests that P2 binding to araI is necessary for RNAP/PBAD complex formation [7,30]. Since the binding of a molecule of RNAP to PBAD would prevent further binding of additional molecules to the same site at the same time, the rate of association is assumed to be inversely proportional to the number of associations. This inverse proportionality represents a feedback loop on the rate of change of the state and is also found in equations (2.4) - (2.8) where

each state is fed back upon itself.

The autoregulation of the C gene is represented in equation (2.4) by the direct proportionality of dx_2/dt to u_2 and the inverse proportionality to x_3 and x_4. Here, the data suggests that both P1 and P2 bound to araO will prevent the binding of RNAP at Pc [15]. Evidence exists supporting stimulation of Pc activity by cAMP(CRP) binding [29]. Thus, a term involving x_6 is also included in equation (2.4). Because of this binding by P1 and P2 to araO, both x_3 and x_4 must feed back on each other as well as upon themselves (equations 2.5 and 2.6). Also present in equations (2.5) and (2.6), are terms relating the rate of binding between P1 and araO to the amount of P1 (x_7) and the rate of binding between P2 and araO to the amount of P2 (x_8).

In equation (2.7), the rate of binding of P2 to araI is dependant upon the amount of P2 available and subject to the repressive effects of araO activity [4]. Evidence indicates that cAMP(CRP) binding also stimulates PBAD activity and a term involving x_6 is placed in equation (2.7) to account for this effect [31,32,33].

The rate of association of the cAMP(CRP) complex to the DNA is directly related to the amount of free cAMP present (u_3) in equation (2.8). Also in equation (2.8), the negative feedback of x_1 upon x_6 is included to represent catabolite repression. That is, the products of L-arabinose catabolism reduce the amount of cAMP available for binding.

The rate of change of the concentration of P1 (equation 2.9) is directly proportional to the production of P1 (araC activity) and inversely proportional to the consumption of P1 by binding at araO (x_3 activity). In equation (2.10), the rate of change of the concentration of P2 is related to the binding of P2 to araO and araI. Coefficients a_{21}, a_{22}, a_{25}, and a_{26} represent the dynamic equilibrium that exists between P1 and P2. Input of L-arabinose shifts this equilibrium toward P2 as represented by the negative coefficient on u_1 in equation (2.9) and the positive coefficient on u_1 in equation (2.10) [17].

Equations (2.3) - (2.11) can be rewritten in state space form as:

$$\dot{X} = \begin{bmatrix} -a_1 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 \\ 0 & -a_3 & -a_4 & -a_5 & 0 & a_6 & 0 & 0 \\ 0 & 0 & -a_7 & -a_8 & 0 & 0 & a_9 & 0 \\ 0 & 0 & -a_{10} & -a_{11} & 0 & 0 & 0 & a_{12} \\ 0 & 0 & -a_{13} & 0 & -a_{14} & a_{15} & 0 & a_{16} \\ -a_{17} & 0 & 0 & 0 & 0 & -a_{18} & 0 & 0 \\ 0 & a_{19} & -a_{20} & 0 & 0 & 0 & -a_{21} & a_{22} \\ 0 & 0 & 0 & -a_{23} & -a_{24} & 0 & a_{25} & -a_{26} \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & b_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & b_2 \\ -b_3 & 0 & 0 \\ b_4 & 0 & 0 \end{bmatrix} U \quad (2.12)$$

$$y = [c_1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \ X \quad (2.13)$$

where:

$$\dot{X} = [\dot{x}_1 \dot{x}_2 \dot{x}_3 \dot{x}_4 \dot{x}_5 \dot{x}_6 \dot{x}_7 \dot{x}_8]^T \quad (2.14)$$

$$X = [x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8]^T \quad (2.15)$$

$$U = [u_1 u_2 u_3]^T \quad (2.16)$$

The following values were then assumed for the coefficients in the A, B and C matrices based in part upon the state relationships mentioned earlier. Unity feedback was assumed for x_1, x_2, x_5 and x_6. Coefficients of .5 were used in pairs to provide unity feedback for x_3 and x_4, and unity consumption of x_8. The catabolite repression of x_6 was taken as 0.5 as the reported value of this repression is 50% [31].

This mathematical model (Figure 2.2) of the L-arabinose gene - enzyme complex was then incorporated into a Fortran program for computer simulation (Figure 2.3). A sixth order Runge-Kutta-Verner routine was utilized for the solution of the eight simultaneous differential equations. All computations were done to five decimal place accuracy. A graphics package was used to provide plots of the response trajectory. The results for the uninduced and induced system are shown in Figures 2.4 and 2.5.

2.2 Parameter Estimation

Utilizing the data of Sheppard [7], the output of the model was compared to the reported values for L-arabinose isomerase specific activity. Sheppard reported an uninduced level of isomerase specific activity of less than or equal to 1.0. Following induction with a 1% arabinose solution (final concentration), a specific activity of 62 was reported. For model comparison purposes, the values of 1.0 and 60.0 were used for the uninduced and induced systems respectively. That is, the expected ratio of specific activity between the induced

and uninduced system was taken as 60:1.

To refine the estimates of the coefficients in the A and B matrices, a least square error criterion was employed. Thus, the model error was defined as:

$$e = (y \text{ expected} - y \text{ computed})^2 \qquad (2.17)$$

$$e_u = (1.0 - y_u)^2 \qquad (2.18)$$

$$e_I = (60.0 - y_I)^2 \qquad (2.19)$$

The selection of values for a_{21}, a_{22}, a_{25}, and a_{26} was arbitrary. No data was found in the literature which could be used to quantify the conversion reaction P1 = P2. Small numbers were chosen because, in the uninduced cell, only a small amount of P2 is believed to be present [15]. These small numbers also made the system relatively insensitive to fluctuations of these values.

The coefficients in the B matrix were adjusted first. Since they are both related to L-arabinose input levels, coefficients b_3 and b_4 were adjusted together in the original system model (Figure 2.2) to reduce the error (e_I) in the induced system. Because negative state values would represent unrealistic physiological situations, an attempt was made to constrain the model to retain positive state values throughout. A value for b_3 and b_4 of 3.65 was selected even though this caused the error in the induced system to remain high (e_I = 3028). This was done because, above this value state x_3 was driven negative. The value for b_2 was adjusted in the uninduced system to 3.5 which yielded an error (e_u) equal to .00001. The value for b_1 was left at 1.0 as

changes to this value increased the error in the uninduced system.

These values for the B matrix parameters represented a starting point for the development of the models of the system. For each model developed, the parameters of the A matrix were optimized and then the parameters of the B matrix were reexamined for optimization. For all of the models, the output values at time = 600 seconds of the model run were used to determine the level of error. This was sufficient to allow the transient behavior of the system to dissappear and permit only steady state values to be measured for goodness of fit.

This starting model (Figure 2.6) yielded the following results upon computer simulation.

$$y_u = .99667 \qquad e_u = .00001$$

$$y_I = 4.03935 \qquad e_I = 3131.$$

Three main approaches were used to further develop the model. One approach was based upon eigenvalue analysis. This method reduced the number of parameters to be optimized from 26 to eight. For the second approach, an unlimited model was developed that allowed the states of the system to assume negative values. Lastly, two models were developed that limited the states of the system to non-negative values. The latter two models differed only in the manner in which the coefficients were optimized.

For all of the models, the initial step was a sensitivity analysis on the parameters of the A matrix. Equal percentage changes were made in each parameter while holding all of the other parameters constant, and the percent changes in the outputs of the uninduced and

induced system were noted. After the sensitivity analysis, the following steps were used to optimize the model.

1) Select the parameters with the largest percentage increases in the induced output (or one that increases the induced output while decreasing the uninduced output)
2) Optimize with respect to e_I. (ie. use the parameters selected in (1) to increase y_I)
3) Continue optimizing all parameters with respect to e_I or e_u as necessary to produce an output ratio of 60:1 between the induced and uninduced system.

2.3 Eigenvalue Model Development: GENE1/2

The eigenvalues of the LTI model represent the natural modes of the system and can be derived from the A matrix of equation 2.1 [24]. The eigenvalues are found by taking the determinant of the matrix (sI-A) and setting it equal to zero.

$$\det (sI-A) = 0 \qquad (2.20)$$

The eigenvalues are the roots of this equation.

The state space model as in Figure 2.6, was transformed into the Real Jordan Form of Figure 2.7. In this form, the parameters of the A matrix are now the eigenvalues of the system. The equations relating

the original states of the system to the transformed states are as follows:

Let $\tilde{X}$ = Transformed state vector

Then $X = T\tilde{X}$ where T= Transformation matrix (2.21)

$\tilde{X} = T^{-1}X$ where T^{-1} = Inverse of T (2.22)

$\dot{\tilde{X}} = [T^{-1}AT]\tilde{X} + [T^{-1}B]U$ (2.23)

$y = [CT]\tilde{X}$ (2.24)

Using this form of the dynamical equations reduced the number of parameters of the A matrix from 26 to eight. This form also has the property that state dependencies are reduced. The transformed states are now dependent only upon their own eigenvalues, and hence, independant of the other states. This eigenvalue model is labeled as GENE1/2 with GENE1 representing the uninduced system and GENE2 representing the induced system. The real and imaginary parts of the three complex eigenvalues were treated separately for the sensitivity analysis and optimization. The sensitivity analysis was done on these eigenvalues and the results are presented in Table 2.1.

In doing the sensitivity analysis, it was noted that the last eigenvalue (-1.0) had no effect on the system output. This fact is also shown in the transformed C matrix which puts no contribution from the eighth state into the output. The sensitivity analysis also indicated that all of the eigenvalues have a greater effect on the uninduced output than on the induced output, which made it difficult to increase the output ratio (about 4:1 initially) to the desired 60:1. The imaginary part of the second eigenvalue was adjusted first

because it had the greatest influence upon the induced output. Results of optimizing EV2(IMM) with respect to e_I are presented in Table 2.2. The other eigenvalues were similarly adjusted, at times with respect to both e_u and e_I in order to increase the output ratio toward 60:1. The results of these adjustments are presented with selected values in Table 2.3. Note that the fourth eigenvalue (REL and IMM parts) was readjusted and the remaining eigenvalues checked for further optimization.

The final results yielded an output ratio of 65:1 with e_u equal to .12004 and e_I equal to 295.

Following the optimization of the eigenvalues, the system was inversely transformed to return to the states of the original system. The resulting new A matrix for the state space model is shown in Figure 2.8.

This model was then run for both non limited and limited state values. That is, the model was run letting the states of the system assume negative values and then run under the condition that the model restricted the states of the system to non negative values. Results of these runs are presented in Table 2.4. The resulting output ratio for the non limited run was 26:1. The difference probably coming from round off error during the computer transformation between the system forms. The optimization of EV4(REL) in Table 2.3 reveals the sensitivity of the model to changes in the third decimal place. This sensitivity could have caused the differences noted between the two state space forms.

The coefficients of the B matrix were then readjusted in an attempt to optimize the system. Table 2.5 shows the result of this optimization.

2.4 Non Limited States Model Development: GENE1C/2C

The second method used in model development involved the formation of a model that allowed the state variables to assume both positive and negative values (GENE1C/2C). This method produced, like the eigenvalue model, a truly linear model. Again, the starting point was the state space model of Figure 2.6.

A sensitivity analysis as described earlier was run on each of the parameters of the A matrix (Table 2.6). Several constants were found that when changed caused an increase in the induced output while causing a decrease in the uninduced output. In addition, the system was found to be very sensitive to changes in parameters a_7 and a_8. A fifty percent increase in a_8 and a fifty percent reduction in a_7 caused the system response to exceed the limits of the computer program. Again, the system showed relative insensitivity to the parameters involved in the P1 to P2 conversion (a_{21}, a_{22}, a_{25} and a_{26}). Hence, these parameters were not changed during the optimization procedure.

In the parameter optimization procedure, the value of the uninduced output was not permitted to fall below zero during optimization of a coefficient with respect to the induced output. All other possible state excursions were permitted. During the

optimization, one parameter (a_4) was found to yeild the smallest e_I when its value was zero.

Optimization of the parameters yielded the values found in Table 2.7, and the following model response.

$y_u = .83017 \quad e_u = .02884$

$y_I = 59.99256 \quad e_I = .00006$

Lastly, the parameters of the B matrix were readjusted as in Table 2.8, yeilding the following improved system response.

$y_u = .99946 \quad e_u = .00000$

$y_I = 60.00013 \quad e_I = .00000$

2.5 Limited Model Development:GENE3A/3B;GENE0A/0B

The need for a limited model arose because the states of the system as defined have no physiological meaning when they are negative. The two previous models, GENE1/2 and Gene1C/2C, can be properly termed linear mathematical models. GENE1/2 was developed based upon eigenvalue analysis and GENE1C/2C was developed based upon coefficient optimization. In both of these models, the state values were allowed to take on negative values in order to best fit the output criteria. In order to produce a model that better fit the physiological situation, two limited models were developed. In both, the states were reset to zero if their calculated value attempted to go negative, and thus their state values were limited to non negative values only. Each state value was calculated at one second intervals by the program.

Two different approaches were used to develop the limited models. In one, GENE3A/3B, the B matrix parameters b_1 and b_2 were readjusted after a sensitivity analysis was performed. The results of their analysis and optimization are shown in Table 2.9, along with the resulting system response. B matrix parameters b_3 and b_4 were left for optimization after adjustments were made to the A matrix parameters.

A sensitivity analysis was then done on the A matrix parameters. The results (Table 2.10), were very similar to those done for the non limited model. This is to be expected as most state values remained positive and were thus unaffected by the state limiter. Some coefficients, though, did differ significantly in their sensitivities between models (a_{12} for example). For optimization, three major classes of coefficients were initially optimized; then, each coefficient was optimized individually. These three classes were those coefficients involved with state feedback, output production and activator concentration. The results of the optimization are presented in Table 2.11.

Lastly, coefficients b_3 (-4.47) and b (+4.47) were optimized resulting in the following system performance.

y_u = 1.00235 $\quad e_u$ = .00001

y_I =60.02403 $\quad e_I$ = .00058

GENEOA/OB, the final model developed, utilized the same sensitivity analysis as GENE3A/3B. In the optimization of this model, an attempt was made to provide satisfactory system performance by

adjusting only selected coefficients. Those coefficients representing state feedback, repression and competition were not adjusted in the development of GENEOA/08. They were held constant in an attempt to make the model physiologically simple and to see what pattern might develop with the remaining coefficients. For this procedure, the coefficients were grouped and optimized as in Table 2.12. Note that only five coefficients were adjusted, along with the B matrix parameters, in order to optimize the system. A significant difference in their optimization was that the adjustments were done based upon the output ratio of 60:1, reported by Sheppard, between the induced and uninduced system, and not upon the actual outputs. Once a ratio of about 60:1 was obtained, the actual outputs were then scaled by changing the value of c_1 in the output matrix. The resulting model yielded the following values.

$y_u = 1.00235 \quad e_u = .00001$

$y_I = 60.02403 \quad e_I = .00058$

$$\dot{X} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & -.5 & -.5 & 0 & 1 & 0 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & 1 & 0 & 1 \\ -.5 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -.001 & .005 \\ 0 & 0 & 0 & -.5 & -.5 & 0 & .001 & -.005 \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} U$$

$$Y = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \ X$$

Original System Model

Figure 2.2

```
// JOB ,
// TIME=(0,30),REGION=2048K
/*JOBPARM LINES=3000
// EXEC PLOTCV77
//SYSLIB DD
//        DD
//        DD
//        DD DSN=SYS1.IMSL.DOUBLE,DISP=SHR
//GO.SOURCE DD *
C    ************************************************************
C    *                                                          *
C    *      GENE-ENZYME SIMULATION                              *
C    *                                                          *
C    *  THIS PROGRAM SIMULATES THE OPERATION OF THE L-          *
C    *ARABINOSE GENE-ENZYME COMPLEX. THE COMPLEX CONSISTS       *
C    *OF THREE STRUCTURAL GENES AND ONE CONTROL GENE. IT        *
C    *IS REGULATED BY BOTH A REPRESSOR (P1) AND AN ACT-         *
C    *IVATOR (P2) PROTEIN AND IS INDUCED BY THE SUGAR L-        *
C    *ARABINOSE. A STATE SPACE REPRESENTATION IS USED TO        *
C    *MATHEMATICALLY SIMULATE THIS GENE-ENZYME COMPLEX.         *
C    *                                                          *
C    *STATE VARIABLES:                                         *
C    *      X(1)=RNAP/PBAD ASSOCIATIONS                        *
C    *      X(2)=RNAP/PC ASSOCIATIONS                          *
C    *      X(3)=P1/ARAO ASSOCIATIONS                          *
C    *      X(4)=P2/ARAO ASSOCIATIONS                          *
C    *      X(5)=P2/ARAI ASSOCIATIONS                          *
C    *      X(6)=CAMP(CAP)/CRP ASSOCIATIONS                    *
C    *      X(7)=CONCENTRATION OF P1                           *
C    *      X(8)=CONCENTRATION OF P2                           *
C    *                                                          *
C    *SYSTEM INPUTS:                                          *
C    *      U1=CONCENTRATION OF L-ARABINOSE                    *
C    *      U2=CONCENTRATION OF RNAP                           *
C    *      U3=CONCENTRATION OF CAMP                           *
C    *                                                          *
C    ************************************************************
C
C     INITIALIZATION
C
C
C   DECLARATIONS
C
          EXTERNAL FCN
          INTEGER N,NW,INIT
          REAL*8 W(50,50)
          REAL*8 C(24)
          REAL*8 X(8),XPRIME(8)
          REAL*8 COST
          REAL TIME(600),ENZY(600)
          REAL*8 T,TEND,Z,Y
          COMMON A
          COST=0.0
          A=1.0
C
C   LOOP FOR COEFFICIENT VARIATION
C
          DO 100 K=1,1,1
C
```

Simulation Program

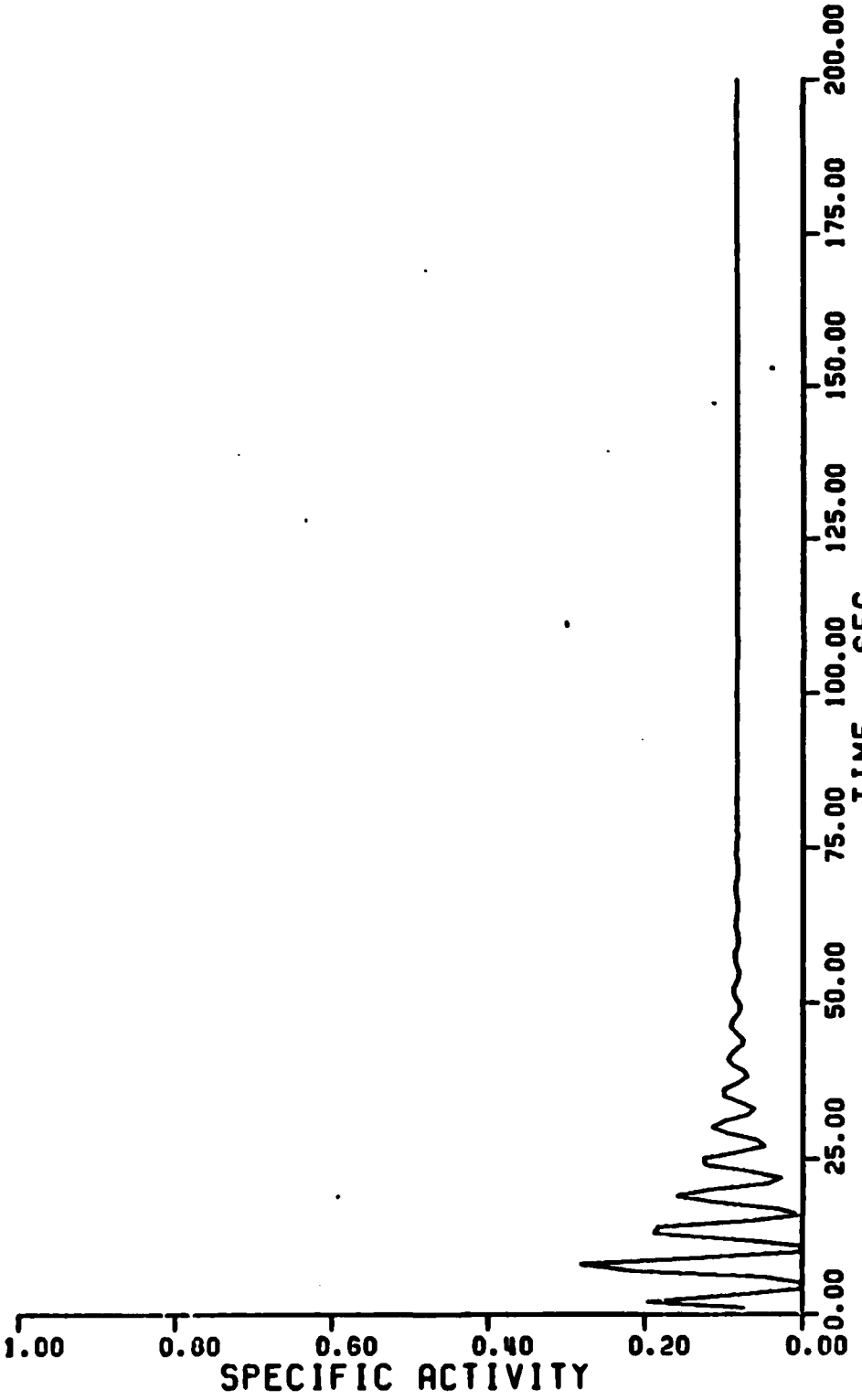
Figure 2.3

Figure 2.3 (Continued)

```
C  SET INITIAL VALUES
C
      T=0.0
      Z=0.0
      Y=0.0
      N=8
      NW=50
C
C     INPUT INITIAL CONDITIONS
C
      X(1)=0.0
      X(2)=0.0
      X(3)=0.0
      X(4)=0.0
      X(5)=0.0
      X(6)=0.0
      X(7)=0.0
      X(8)=0.0
C
C  SET PRINT HEADINGS
C
      PRINT 61
 61   FORMAT ('0',5X,'X1',10X,'X2',10X,'X3',10X,'X4',10X,'X5')
      PRINT 62
 62   FORMAT ('+',66X,'X6',10X,'X7',10X,'X8',8X,'T',10X,'Y',9X,'A')
      INIT=1
C
C  SET LOOP FOR SIMULATION
C
      DO 90  I=1,600,1
        TEND=T+1.0
        TIME(I)=TEND
        CALL DVERK (N,FCN, T,X,TEND,.05,INIT,C,NW,W,IER)
        DO 66 J=1,8
          IF (X(J).LT.0.0) THEN
          X(J)=0.0
          END IF
 66     CONTINUE
        T=TEND
        Y= 37.4672*X(1)
        ENZY(I)=Y
        PRINT 80,X,TEND,Y,A
 80     FORMAT (' ',8(F10.5,2X),F7.2,F15.5,2X,F6.3)
 90   CONTINUE
C
C  CALCULATE COST FUNCTION
C
      COST=X(1)**2+X(2)**2+X(3)**2+X(4)**2+X(5)**2+X(6)**2+
     C   X(7)**2+X(8)**2
      PRINT 95,COST
 95   FORMAT ('1',10X,'COST = ',F15.5)
      A=A+.1
100   CONTINUE
C
C  GRAPHICS ROUTINES
C
      CALL FACTOR (.9)
      CALL NEWPEN (3)
      CALL AXIS (1.5,1.5,'SPECIFIC ACTIVITY',-17,5.,0.,5.,-1.,1.)
      CALL AXIS (6.5,1.5,'TIME - SEC',-10,0.,90.,0.,25.,1.)
```

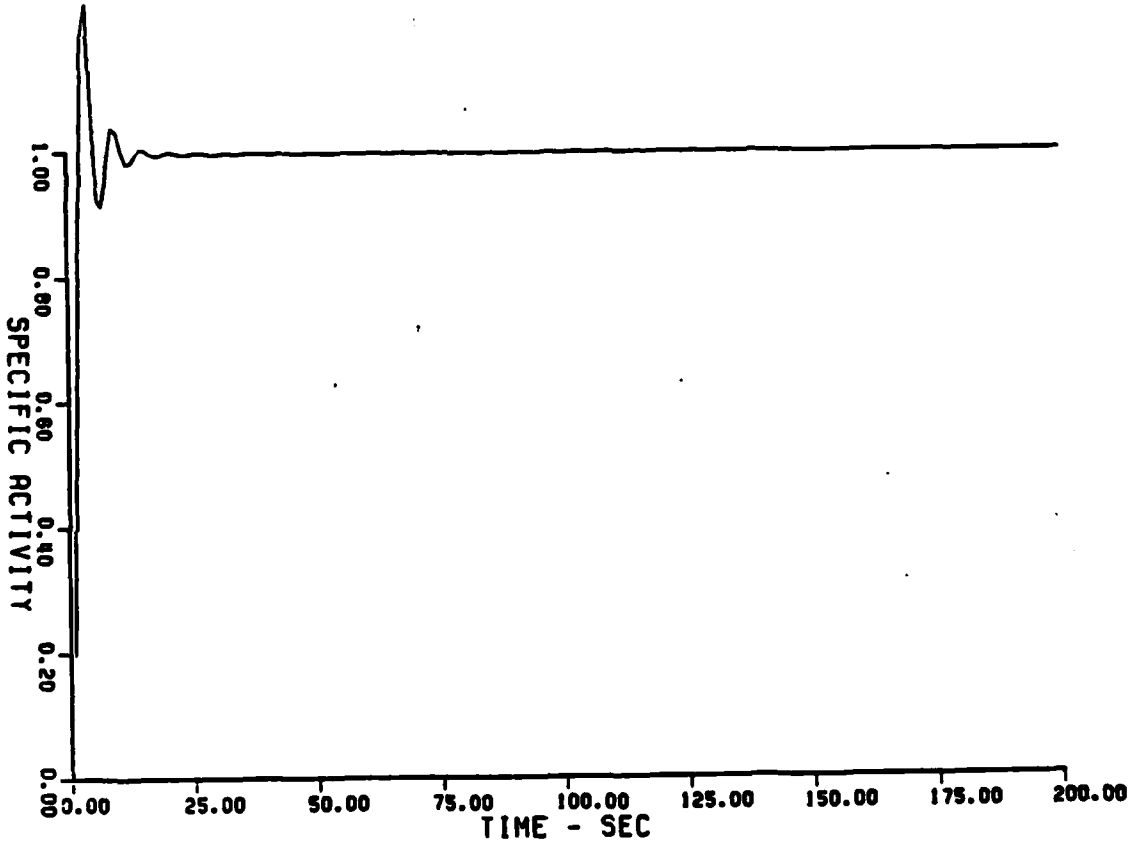
Figure 2.3 (Continued)

```
      CALL NEWPEN (2)
      CALL LINE (ENZY,6.5,-1.,TIME,-37.5,25.,200,0,0)
      CALL PLOTE2
      STOP
      END
C
C     SUBROUTINE FOR DIFFERENTIAL EQUATION SOLUTION
C
      SUBROUTINE FCN (N,T,X,XPRIME)
        REAL*8 X(N),XPRIME(N)
        REAL*8 T,TEND
        COMMON A
        INTEGER N
        XPRIME(1)=-1.0*X(1)+1.0*X(5)
        XPRIME(2)=-1.0*X(2)-.5*X(3)-.5*X(4)+1.0*X(6)+1.0
        XPRIME(3)=-.5*X(3)-.5*X(4)+1.0*X(7)
        XPRIME(4)=-.5*X(3)-.5*X(4)+1.0*X(8)
        XPRIME(5)=-X(3)-X(5)+1.0*X(6)+1.0*X(8)
        XPRIME(6)=-.5*X(1)-X(6)+1.0
        XPRIME(7)=X(2)-X(3)-.001*X(7)+.005*X(8)
        XPRIME(8)=-.5*X(4)-.5*X(5)+.001*X(7)-.005*X(8)
      RETURN
      END
/*
//
```



Uninduced Response/Original Model

Figure 2.4



Induced Response/Original Model
Figure 2.5

$$\dot{X} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & -.5 & -.5 & 0 & 1 & 0 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & 1 & 0 & 1 \\ -.5 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -.001 & .005 \\ 0 & 0 & 0 & -.5 & -.5 & 0 & .001 & -.005 \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3.5 \\ -3.65 & 0 & 0 \\ 3.65 & 0 & 0 \end{bmatrix} U$$

$$Y = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0] X$$

Starting System Model

Figure 2.6

$$\dot{\tilde{x}} = \begin{bmatrix} -1.909 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & -0.067 & 1.067 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & -1.067 & -0.067 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & -0.701 & 0.601 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & -0.601 & -0.701 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.331 & 0.361 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.861 & -0.331 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 \end{bmatrix} \tilde{x} + \begin{bmatrix} -0.265 & 0.134 & -1.753 \\ -2.345 & 0.672 & 1.270 \\ 3.426 & -0.225 & 1.083 \\ -1.499 & -0.662 & 3.338 \\ -1.128 & -0.338 & 0.596 \\ -0.967 & -0.338 & 1.507 \\ 1.258 & -0.921 & -1.647 \\ 0.000 & -0.715 & 0.000 \end{bmatrix} u$$

$$y = \begin{bmatrix} -1.000 & 0.617 & -0.342 & -0.256 & -0.967 & 0.199 & 0.553 & 0.000 \end{bmatrix} \tilde{x}$$

Real Jordan Normal Form of Starting Model

Figure 2.7

$$\begin{bmatrix}
-0.536 & -0.546 & 0.220 & 0.068 & 0.541 & 0.703 & -0.101 & -0.489 \\
-0.315 & -0.794 & -0.058 & -0.262 & 0.413 & -0.069 & 0.191 & 0.127 \\
-0.131 & -0.354 & -0.204 & -0.140 & 0.039 & -0.373 & 0.354 & 0.337 \\
-0.035 & -0.307 & -0.033 & -0.296 & 0.296 & 0.054 & -0.073 & -0.006 \\
-0.351 & -0.101 & -0.091 & 0.442 & -0.292 & -0.006 & 0.119 & 0.122 \\
-0.048 & -0.139 & -0.223 & 0.014 & -0.212 & -0.536 & 0.273 & 0.444 \\
0.234 & 0.378 & -0.162 & -0.037 & -0.171 & -0.097 & -0.063 & 0.065 \\
-0.079 & -0.037 & -0.024 & 0.102 & -0.161 & -0.229 & 0.108 & 0.166
\end{bmatrix}$$

New A Matrix for Eigenvalue Model

Figure 2.8

Table 2.1

GENE1/2

Eigenvalue Sensitivity

Eigenvalue	Value	%Change	Y_U	Y_I	$\%Y_U$	$\%Y_I$
EV1	-2.352	+30	.7810	3.79748	20.8	6.0
(-1.809)	-1.266	-30	1.37858	4.48462	38.6	11.1
EV2(REL)	-.047	+30	.97513	4.06638	1.97	.7
(-.067)	-.087	-30	1.01326	4.00789	1.86	.7
EV2(IMM)	1.387	+30	1.25636	3.48932	26.3	13.6
(1.067)	.747	-30	2.21534	5.02592	122.7	24.5
EV3(REL)	-.911	+30	.54316	3.73432	45.4	7.5
(-.701)	-.491	-30	1.73999	4.43286	74.9	9.8
EV3(IMM)	.781	+30	1.17891	3.94339	18.5	2.3
(.601)	.421	-30	.59353	4.05764	40.3	.5
EV4(REL)	-.430	+30	.97184	3.99263	2.3	1.1
(-.331)	-.232	-30	1.07783	4.11185	8.4	1.8
EV4(IMM)	1.119	+30	1.56751	4.27768	57.6	5.9
(.861)	.603	-30	-.00945	3.58947	101	11.1

Table 2.2

GENE1/2

Eigenvalue Optimization (EV3)

Parameter (EV3)	e_U	e_I	Y_U	Y_I
1.067	.00003	3131.	.99472	4.03792
.067		2324.		11.78793
.077		2210.		12.97969
.087		2145.		13.67804
.097		2113.		14.02753
.150		2179.		13.31588
.120		2112.		14.04074
.110		2103.		14.03859
.107		2103.52		14.13581
.109		2103.52		14.13580
.108	175.	2103.45	14.25475	14.13658

Table 2.3

GENE1/2

Eigenvalue Optimization

Parameter		e_U	e_I	Y_U	Y_I
EV3(IMM)	.5	170.	2101.	14.06000	14.16166
	.45	167.	2101.	13.93681	14.16138
	.40	163.	2102.	13.79155	14.15044
EV2(REL)	-.005	157.	1160.	13.53817	25.92913
	-.010	150.	1201.	13.25876	25.34309
	-.015	159.	1257.	13.61296	24.53841
EV4(IMM)	.06	64.			
	.10	61.9	1379.	8.86801	22.85656
	.16	63.			
EV4(REL)	-.135	.02650		.83721	
	-.137	.00043	1635.	.97936	19.56301
	-.140	.03613		1.19007	
EV1	-.04		31.		65.62150
	-.045	1229.	.15070	36.06218	60.38821
	-.06		101.		49.92154
EV4(IMM)	.005	62.		8.89327	
	.019	29.	215.	6.40233	45.31187
	.020	29.		6.47619	
EV4(REL)	-.03		387.	-18.68295	
	-.044	.12004	295.	.65353	42.80333
	-.045	.48675	280.	1.69767	43.25549

Table 2.4

GENE1/2

Eigenvalue Model

	Non Limited Run		Limited Run	
	Uninduced	Induced	Uninduced	Induced
X1	1.68092	44.32712	18.58793	34.59979
X2	-20.28888	-22.09476	0	0
X3	1.54579	-32.72974	6.33231	0
X4	-46.65178	13.51421	0	34.17903
X5	-67.90229	-9.58770	0	30.51621
X6	27.10089	31.80638	16.59802	33.68437
X7	23.58632	21.07048	28.41276	0
X8	-26.01703	-6.01371	0	49.35741
Y	1.68092	44.32712	18.58793	34.59979
e	.46365	245.	309.	645.

Table 2.5

GENE1/2

B Parameter Readjustment

Parameter		eu	ei	Yu	Yi
b1	1.03	.02646		1.16266	
	1.04	.00010	267.	.98991	43.63606
	1.05	.03343		.81716	
b2	3.49	.00414		.93568	
	3.50	.00010	267.	.98984	43.63606
	3.51	.00194		1.04400	
b3&	5.04		.01522		59.87664
b4	5.05		.00004		59.99347
	5.06		.01217		60.11030

Table 2.6

GENE1C/2C

Parameter Sensitivity

Parameter		%Change	Yu	Yi	%Yu	%Yi
a1	-.5	-50	1.70898	6.92624	71.5	71.5
	-1.5	+50	.70346	2.85102	29.4	29.4
a2	+.5	-50	.54355	2.20295	45.5	45.5
	+1.5	+50	1.38018	5.59366	38.5	38.5
a3	-.5	-50	.61821	3.81436	37.9	5.6
	-1.5	+50	1.18577	4.15177	18.9	2.8
a4	-.25	-50	.84505	4.13333	15.2	2.3
	-.75	+50	1.10481	3.97226	10.8	1.7
a5	-.25	-50	1.04059	3.89672	4.4	3.5
	-.75	+50	.95612	4.17070	4.0	3.3
a6	+.5	-50	1.23872	4.15872	24.3	3.0
	+1.5	+50	.73338	3.90958	26.4	3.2
a7	-.25	-50	------	------	---	---
	-.75	+50	.99729	4.03896	.06	0
	-.40	-20	.99641	4.03950	.03	0
	-.35	-30	.68847	4.15257	31.0	2.8
	-.3375	-32.5	-6.90601	7.29134	793	80.5
a8	-.25	-50	.99687	4.03867	.02	0
	-.75	+50	------	------	---	---
	-.675	+35	2.07047	4.76635	107.7	18.0
	-.70	+40	129.73193	53.70491	13016	1329

Table 2.6 (Continued)

Parameter	%Change	Yu	Yi	%Yu	%Yi
a9 +.5	-50	.99750	4.03993	.08	.01
+1.5	+50	.99639	4.03915	.02	0
a10 -.25	-50	.62400	4.27035	37.4	5.7
-.75	+50	1.37033	3.80772	37.5	5.7
a11 -.25	-50	1.13942	3.57682	14.3	11.5
-.75	+50	.88590	4.39824	11.1	8.9
a12 +.5	-50	1.39424	4.31599	39.9	6.8
+1.5	+50	-.42201	1.73232	142	57.1
a13 -.5	-50	1.74875	3.57314	75.5	11.5
-1.5	+50	.24358	4.50617	75.6	11.5
a14 -.5	-50	1.32976	5.38874	33.4	33.4
-1.5	+50	.79707	3.23040	20.0	20.0
a15 +.5	-50	.13741	3.61559	86.2	10.5
+1.5	+50	1.66469	4.36880	67.0	8.2
a16 +.5	-50	.71037	3.84013	28.7	4.9
+1.5	+50	1.21932	4.19428	22.3	3.8
a17 -.25	-50	1.08711	4.40589	9.0	9.0
-.75	+50	.92012	3.72911	7.7	7.7
a18 -.5	-50	1.85306	4.46169	85.9	10.5
-1.5	+50	.64415	3.86550	35.4	4.3
a19 +.5	-50	1.29920	4.21920	30.4	4.5
+1.5	+50	.78045	3.91081	21.7	3.2

Table 2.6 (Continued)

Parameter		%Change	Yu	Yi	%Yu	%Yi
a20	-.5	-50	.61723	4.27455	38.1	5.8
	-1.5	+50	1.18591	3.92204	19.0	2.9
a21	-.01	+900	.99817	4.04039	.15	.03
	-.005	+40	.99734	4.03981	.07	.01
a22	+.001	-80	.99734	4.03981	.07	.01
	+.01	+100	.99582	4.03876	.09	.01
a23	-.25	-50	.74606	4.85129	25.1	20.1
	-.75	+50	1.12208	3.63302	12.6	10.1
a24	-.25	-50	1.19560	4.84560	20.0	20.0
	-.75	-50	.85449	3.46312	14.3	14.3
a25	+.005	+40	.99933	4.04120	.27	.05
	+.01	+900	1.00268	4.04350	.60	.10
a26	-.001	-80	.99933	4.04120	.27	.05
	-.01	+100	.99334	4.03703	.33	.06

Table 2.7

GENE1C/2C

Parameter Optimization

Parameter		e_u	e_i	Yu	Yi
a13	-1.65	.96539		.01745	
	-1.66	.99525	3062.	.00238	4.65569
	-1.67	1.02551		-.01270	
a1	-.47	.47634		1.69017	
	-.48	.00609	2482.	1.07802	10.17830
	-.49	.12689		.64378	
a2	+1.09		2401.		10.99756
	+1.10	.39057	2350.	.37505	11.52105
	+1.11		2374.		11.27536
a3	-1.58	.00006		.99219	
	-1.59	.00001	2501.	1.00322	9.98774
	-1.60	.00020		1.01416	
a4	-.01	.58430	2361.	.23560	11.40848
	0.0	.61798	2357.	.21388	11.44856
a5	-1.37		2235.		12.72093
	-1.38	.63843	2233.	.20098	12.73570
	-1.39		2244.		12.62639
a6	+.97	.65696		.18947	
	+.98	.59221	2220.	.23045	12.87642
	+.99	.60006		.22536	

Table 2.7 (Continued)

Parameter		eu	ei	Yu	Yi
a7	-.49	1.58304	2127.	-.25819	13.87986
	-.50	.59221	2220.	.23045	12.87642
	-.51		2239.		12.67817
a8	-.48	.77055		.12219	
	-.49	.39472	2259.	.37174	12.46951
	-.50	.59221		.23045	
a9	+1.10	.36549		.39544	
	+1.11	.27693	2269.	.47376	12.36558
	+1.12	.27707		.47363	
a10	-.48	1.0638		-.03141	
	-.49	.65532	2212.	.19048	12.96163
	-.50	.27693	2269.	.47376	12.36558
a12	+.53	.00025		1.01573	
	+.54	.00019	2363.	.98632	11.37932
	+.55	.00179		.95773	
a17	-.06		2155.		13.57247
	-.07	.02124	2149.	1.14575	13.63696
	-.08		2156.		13.56213
a23	-.05		1902.		16.38507
	-.02	.93986	1875.	.03054	16.68826
	-.01	1.01867	1866.	-.00929	16.79495

Table 2.7 (Continued)

Parameter		eu	ei	Yu	Yi
a24	-.15		58.		52.34731
	-.13	.79301	.13846	.10949	59.62790
	-.12		16.		64.08450
a19	+.19	.00000	13.28524	1.00203	56.35511
a20	-.35		.01759		59.86731
	-.34	.22992	.00001	.52051	60.00359
	-.33		.02139		60.14625
a18	-.57	.00061		.97527	
	-.56	.00004	.00849	.99398	59.90786
	-.55	.00018		1.01335	
a14	-.99	.00002	.00054	.99512	59.97673
a15	+.94		.00001		59.99666
	+.93	.01419	.00000	.88087	59.99999
	+.92		.00001		60.00332
a11	-.40	.00018		.98676	
	-.39	.00000	.40575	.99909	59.36302
	-.38	.00014		1.01180	
a16	+.61		.00012		60.01079
	+.62	.02884	.00006	.83017	59.99256
	+.63		.00065		59.97441

Table 2.8

GENE1C/2C

B Matrix Optimization

Parameter		eu	ei	Yu	Yi
b2	4.1	.00001	.02571	.99797	60.16035
b1	.99	.00000	.02619	.99946	60.16184
b3	-3.52		.00000		60.00013

Table 2.9

GENE3A/3B

B Matrix Parameter Sensitivity

Nominal Values: Y_U = .51989 Y_I = 3.82788

Parameter		%Change	Yu	Yi	%Yu	%Yi
b_1	1.5	+50	.47495	3.81210	8.6	.4
	.5	-50	.56484	3.84365	8.6	.4
b_2	1.75	-50	.21500	3.04453	58.6	20.5
	5.25	+50	.82478	4.60902	58.6	20.4
b_3	-1.825	-50		3.73621		2.4
	-5.475	+50		3.88546		1.5
b_4	1.825	-50		2.73688		28.5
	5.475	+50		4.91888		28.5

Parameter Optimization

Parameter		eu	ei	Yu	Yi
b_2	6.24	.00001		.99726	
	6.25	.00000		.99901	
	6.26	.00000	3029.	1.00075	4.95793
b_1	.99	.00000		1.00165	
	1.01	.00000	3029.	.99985	4.95621
	1.02	.00000		.99895	

System Performance

y_U = .99985 e_u = .00000

y_I = 4.95621 e_I = 3029.

Table 2.10

GENE3A/3B

Parameter Sensitivity

Nominal Values: Y_U = .99985 Y_I = 4.95621

Parameter		%Change	Yu	Yi	%Yu	%Yi
a_1	-.5	-50	1.83923	8.26798	84.0	66.8
	-1.5	+50	.68670	3.49810	31.3	29.4
a_2	+.5	-50	.52268	2.70293	47.7	45.5
	+1.5	+50	1.43720	6.82835	43.7	37.8
a_3	-.5	-50	.67290	4.43821	32.7	9.2
	-1.5	+50	1.70667	5.10071	70.7	2.9
a_4	-.25	-50	.86906	4.95697	13.1	0
	-.75	+50	1.36739	4.95570	36.8	0
a_5	-.25	-50	1.02260	4.85385	2.3	2.1
	-.75	+50	.97804	5.05247	2.2	1.9
a_6	+.5	-50	2.13708	5.11794	113.7	3.3
	+1.5	+50	.73502	4.63918	26.5	6.4
a_7	-.25	-50	1.12003	4.95620	12.0	0
	-.75	+50	1.00151	4.95621	.2	0
a_8	-.25	-50	.99976	4.95572	0	0
	-.75	+50	.99996	4.95670	0	0
a_9	+.5	-50	1.00244	4.95716	.3	0
	+1.5	+50	.99904	4.95589	.1	0
a_{10}	-.25	-50	.93620	4.95792	6.4	0
	-.75	+50	1.31124	4.95450	31.1	0

Table 2.10 (Continued)

Parameter	%Change	Yu	Yi	%Yu	%Yi
a_{11} -.25	-50	1.04713	4.62450	4.7	6.7
-.75	+50	.95525	5.21359	4.5	5.2
a_{12} +.5	-50	1.01289	5.41736	1.3	9.3
+1.5	+50	.99104	4.74667	.9	4.2
a_{13} -.5	-50	2.42318	4.95271	142.4	.1
-1.5	+50	.35914	4.95971	64.1	.1
a_{14} -.5	-50	1.35708	6.61192	35.7	33.4
-1.5	+50	.88336	3.96367	11.7	20.0
a_{15} +.5	-50	.18543	3.87358	81.5	21.8
+1.5	+50	2.61971	5.79790	162.0	17.0
a_{16} +.5	-50	.98461	4.62410	1.5	6.7
+1.5	+50	1.00768	5.21450	.8	5.2
a_{17} -.25	-50	1.04536	5.40587	4.6	9.1
-.75	+50	.95813	4.55223	4.2	8.2
a_{18} -.5	-50	2.02729	6.03514	102.8	21.8
-1.5	+50	.65533	4.45965	34.5	10.0
a_{19} +.5	-50	2.14731	5.12187	114.8	3.3
+1.5	+50	.81323	4.69451	18.7	5.3
a_{20} -.5	-50	.67256	4.95799	32.7	0
-1.5	+50	1.70623	4.95532	70.6	0
a_{23} -.25	-50	.81060	5.54144	18.9	11.8
-.75	+50	1.27538	4.66471	27.6	5.9
a_{24} -.25	-50	1.47985	5.94741	48.0	20.0
-.75	-50	.80441	4.24912	19.5	14.3

Table 2.11

GENE3A/3B

Parameter Optimization

State Feedback Coefficients (SFC)
(a_1, a_3, a_7, a_8, a_{10}, a_{11}, a_{14}, a_{18})
Output Production Coefficients (OPC)
(a_2, a_{16})
Activator Concentration Coefficients (ACC)
(a_{23}, a_{24})

Parameter		e_U	e_I	Y_U	Y_I
SFC	-.4	.16165	2295.	1.40206	12.08762
OPC	2.5	1.68605	1497.	2.29848	21.30729
ACC	-.13	1.00000	3.12039	0.00000	61.76646
a_5	-.34	.74680	1498.	1.86418	21.28918
a_4	-.32	.01259	1498.	1.11219	21.29160
a_9	.3	1.19399	1.92534	2.09270	61.38756
a_{17}	-.65	.05643	50.02587	1.23765	52.92710
a_{16}	2.1	.02442	127.	.84372	48.72525
a_6	.9	.00510	127.	1.07141	48.72179

Table 2.12

GENEOA/OB

Coefficient Classes with Optimized Values

State Feedback	Repression	Competition
$a_1 = -1.0$	$a_4 = -.5$	$a_{23} = -.5$
$a_3 = -1.0$	$a_5 = -.5$	$a_{24} = -.5$
$a_{14} = -1.0$	$a_{13} = -1.0$	
$a_{18} = -1.0$		Production
$a_7 = -.5$	Catabolite Repression	$a_{19} = 1.0$
$a_8 = -.5$	$a_{17} = -.5$	$a_2 = 1.0$
$a_{10} = -.5$		
$a_{11} = -.5$		

Variable Dependencies

$a_2 = 1.0$

$a_6 = .1$

$a_9 = .9$

$a_{12} = 2.5$

$a_{15} = .4$

$a_{16} = 5.2$

$a_{19} = 1.0$

Chapter 3

Model Discrimination

3.1 Introduction

Four models of the L-arabinose system were developed based upon state space representation. It was decided to use only one of the models for further experimentation. Extrinsic parameters were used to discriminate among the models and select the model of best fit. In extrinsic parameter discrimination one uses criteria, other than those used for model development to test the performance of each model [8]. The model that best satisfies the criteria is then chosen as the model of best fit. The following parameters were selected as the criteria for discrimination.

1) Removal of catabolite repression in the induced system.

2) Induced response to a 0.4% arabinose input.

3) Induced response to one tenth the normal cAMP input.

4) The ratio in the induced system of x_1 to x_2.

5) The induced response when the cAMP input is first eliminated then increased to twice its normal level.

6) The response of the uninduced system when x_2 is set to zero.

7) The ten percent settling time in the induced system.

8) The difference between x_4 and x_5 in the induced system.

9) The number of negative or zero state values in the uninduced and

induced system.

These criteria include both input/output performance measures and physiologically specified state relationships. Thus, both the mathematical behavior of the model and the physiological similarity of the model were considered in the discrimination procedure.

3.11 Removal of Catabolite Repression in the Induced System

Catabolite repression, the repressive effect of the products of the enzyme degradation of L-arabinose, was simulated by an x_1 term in the equation for x_6 (equation 2.8). Removal of catabolite repression was accomplished by simply deleting this term. In the literature, it is reported that the induced system response doubles when catabolite repression is removed [31]. The expected result for the induced output was therefore taken as twice the normal induced output.

3.12 Induced Response to a 0.4% Arabinose Input

Sheppard's data, which was derived from bacteria growing in a 1% arabinose media, was used for model development [7]. The data for bacteria grown in a 0.4% arabinose media was obtained from work reported by Katz [31]. In this data, the induced system output is given for three separate bacterial cultures. The average value, 24.5, was taken as the expected value for the models. To simulate a 0.4% arabinose input in the model, the B matrix parameters, b_3 and b_4, were reduced to 40% of their normal value.

3.13 Induced Response to One Tenth the Normal cAMP Input

Decreasing the amount of cAMP added to an induced culture causes a decrease in L-arabinose isomerase production [33]. Data suggests that a 90% reduction in the concentration of added cAMP causes a 60% reduction in the induced output of enzyme [33]. Thus, for an input of one tenth of the normal amount of cAMP, with all other parameters held constant, the expected value of the output was taken as 40% of the normal induced output.

3.14 The x_1:x_2 Ratio in the Induced System

Casadaban [29] reports that in the normal induced system the level of Pc activity is only about 3% that of PBAD activity. This translates into a ratio of 33.3:1 for x_1:x_2 activity in the models.

3.15 Induced Response to Zero Then 2x cAMP Input

Another study using various cAMP inputs yields another parameter for model discrimination. In this study, the amount of cAMP added to the bacterial culture was reduced to zero and then increased to twice the normal input [31]. The bacteria responded with an output 2.28 times as great in the 2x cAMP input case than in the zero cAMP input case. Each model was similarly tested. The output value, produced when no cAMP was input to the model, was multiplied by 2.28 to yield the expected value of the model when twice the normal cAMP was input to

the system. The changes to the cAMP inputs were accomplished by changing coefficient b_2 in the input matrix.

3.16 Uninduced Response When x_2=0

The setting of state x_2 equal to zero (ie. Let x_2=0 in the system equation 2.4) simulates a non functional C gene in the L-arabinose system. The literature reports that a non functional C gene in the uninduced culture results in approximately normal uninduced output [7]. Therefore, a value of 1.0 was assumed for the expected output when x_2 was set to zero in the uninduced system.

3.17 The Ten Percent Settling Time in the Induced System

Experience with the L-arabinose system has shown that the production of L-arabinose isomerase proceeds at a constant rate [31]. Thus, the model in the induced state should have transient behavior that dies out rapidly yielding a steady state production of the enzyme. The earliest sample time for data collection reported was five minutes after the onset of induction. Thus, the expected value for the 10% settling time was taken as 300 seconds. A 10% settling time, the time it takes the output to settle to and remain within 10% of its steady state value, was chosen because this amount of variation is possible in the experimental measurements [31].

3.18 The x_4/x_5 Difference in the Induced System

A recent study has suggested that the activator, P2, shows no preference in binding to either araI or araO [16,19]. Thus states x_4 and x_5 could be expected to have equal values. Hence, the difference between x_4 and x_5 in the steady state of the induced gene can be expected to be zero.

3.19 Number of Negative or Zero State Values ; Induced and Uninduced

Lastly, the definition of the system's states suggests that there should be no negative state values as those would represent a negative activity or negative concentration. Similarly, zero values in those models with a state limiter would represent negative states and therefore violate the state definition. Thus, in comparing the models, the number of negative or zero states was considered as a measure of the deviation from the expected value of zero in both the uninduced and induced systems.

3.2 Evaluation of the Models

Selected characteristics for each of the models developed are listed in Tables 3.1,3.2,3.3 and 3.4. Steady state values are presented for (a) the uninduced system, (b) the normal induced system (1% arabinose) and (c) the system induced with 0.4% arabinose. Also found in these tables are selected transient behavior characteristics for the model. Note that all models eventually settle to steady state operation except GENE3A/3B. The oscillatory behavior of this model was

not detected until model development was complete.

In evaluating the fit of the models to the expected values of the extrinsic parameters, the method of residual analysis was used [8]. A residual can be defined as follows.

Residual = |Expected value - Computed value|

In some cases, the actual residual may not yield as much information as the percent error which can be defined as follows.

Percent error = (Residual/Expected value) x 100

The residual analyses of the four models are presented in Tables 3.5, 3.6, 3.7 and 3.8 with a comparison among the four models presented in Table 3.9.

Examination of Table 3.9 reveals that GENEOA/OB was the best fit model in five out of the nine areas while GENE1/2 was the best fit model for the other four. If each model is scored, for each of the criterion, on a scale of 4 to 1 with 4 representing the best fit and 1 representing the worst fit, the following total scores are obtained for the four models.

GENEOA/OB	28
GENE1/2	24
GENE1C/2C	25
GENE3A/3B	17

From these comparisons, it appears that GENEOA/OB and GENE1/2 yield the best fit to the extrinsic parameters. GENE1/2 is the only model to do reasonably well for the input/output relationships involving cAMP factors (w/o catabolite repression, 0.1 cAMP input and

0/2x cAMP input). The difference in the residuals for a 0.4% arabinose input is well within experimental error for all models.

Overall, GENEOA/OB (Figure 3.1) was judged the model of best fit based upon the preceeding analysis. Although GENE1/2 was superior for some of the criteria, and scored a close second overall, this model was not further evaluated. An additional consideration influenced this decision. The new state dependencies, as manifested in the new A matrix of the system following eigenvalue optimization, are difficult to explain physiologically. Furthermore, this model would be difficult to use in the cost analysis as it would be difficult to simulate genetic mutational conditions with this model.

$$\dot{X} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & -.5 & -.5 & 0 & .1 & 0 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & .9 & 0 \\ 0 & 0 & -.5 & -.5 & 0 & 0 & 0 & 2.5 \\ 0 & 0 & -1 & 0 & -1 & .4 & 0 & 5.2 \\ -.5 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -.001 & .005 \\ 0 & 0 & 0 & -.5 & -.5 & 0 & .001 & -.005 \end{bmatrix} X + \begin{bmatrix} 0 & 0 & 0 \\ 0 & .7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & .36 \\ -.4 & 0 & 0 \\ 1.6 & 0 & 0 \end{bmatrix} U$$

$$Y = [37.4672 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0] X$$

GENEOA/OB

Figure 3.1

Table 3.1

GENEOA/OB

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.02669	1.59421	.66104
X2	.50077	.07694	.31576
X3	.50096	0	.15610
X4	0	1.63394	.61827
X5	.02712	1.63869	.66104
X6	.34659	0	.02948
X7	.26215	.04654	.43021
X8	.07663	.25099	.15488
Y	1.00017	59.73076	24.76729

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	49 sec.	10 sec.
5% Settling Time	35 sec.	7 sec.
10% Settling Time	31 sec.	5 sec.
Peak Overshoot	203%	30.5%
Frequency	.17 Hz	.33 Hz
Time Constant	15 sec.	2 sec.

Table 3.2

GENE1/2

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.94476	60.01078	24.57117
X2	-19.33885	-22.36168	-20.84799
X3	2.02655	-45.37633	-16.93461
X4	-47.28985	35.94694	-13.99513
X5	-68.12986	12.52118	-35.86944
X6	26.27311	33.74719	29.86274
X7	23.57445	20.22070	22.23295
X8	-25.58386	1.96575	-14.56401
Y	.94476	60.01078	24.57117

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	753 sec.	416 sec.
5% Settling Time	581 sec.	244 sec.
10% Settling Time	521 sec.	183 sec.
Peak Overshoot	1951%	29.7%
Frequency	.017 Hz	.017 Hz
Time Constant	61 sec.	142 sec.

Table 3.3

GENE1C/2C

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.99946	60.00013	24.59973
X2	7.83685	-12.53191	-.31066
X3	4.40249	-17.42540	-4.32867
X4	-3.20147	15.02951	4.09092
X5	.43220	25.94601	10.63772
X6	7.19650	-.17859	4.24646
X7	.56984	-1.21463	-.14395
X8	1.68268	-4.95729	-.97331
Y	.99946	60.00013	24.59973

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	109 sec.	66 sec.
5% Settling Time	87 sec.	45 sec.
10% Settling Time	77 sec.	34 sec.
Peak Overshoot	505%	67.8%
Frequency	.048 Hz	.048 Hz
Time Constant	17 sec.	17 sec.

Table 3.4

GENE3A/3B

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	5.67836 *	60.02403	27.53868
X2	5.05832	0	0
X3	8.66590	0	0
X4	0	24.78320	9.34848
X5	1.38540	8.10982	3.89071
X6	4.58111	0	0
X7	9.45382	0	0
X8	1.17530	6.50888	2.45720
Y	5.67837	60.02403	27.53868

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	@	21 sec.
5% Settling Time	@	12 sec.
10% Settling Time	@	11 sec.
Peak Overshoot	294%	12%
Frequency	.17 Hz	.05 Hz
Time Constant	@	11 sec.

* Average value taken over last period

@ Oscillations insufficiently dampped

Table 3.5

GENEOA/OB Residual Analysis

Criterion	Expected	Computed	Residue	%Error
w/o Cat.Rep.	119.46152	63.60912	55.8524	46.75
0.4% ara	24.5	24.76729	.26729	1.09
0.1 cAMP	23.8923	58.86434	34.972	146.7
X1:X2 (I)	33.3:1	20.7:1	12.6	37.84
0/2x cAMP	133.9913	60.69341	73.2979	54.70
X2=0 (u)	1.0	2.15432	1.15432	115.4
10% S.T. (I)	300 sec.	5 sec	295 sec.	
X4/X5 Diff.	0	.0047	.00475	
#-/0 values	0	3	3	

Table 3.6

GENE1/2 Residual Analysis

Criterion	Expected	Computed	Residual	%Error
w/o Cat.Rep.	120.02156	81.16317	38.85839	32.38
0.4% ara	24.5	24.57117	.07117	.29
0.1 cAMP	23.00431	42.97822	18.97391	82.48
X1:X2 (I)	33.3:1	60:1 *	26.7 *	80.18
0/2x cAMP	93.6754	78.93585	14.73957	15.73
X2=0 (u)	1.0	@	@	@
10% S.T. (I)	300 sec.	183 sec.	156 sec.	
X4/X5 Diff.	0	23.42576	23.42576	
#-/0 values	0	6	6	

* X2 negative, taken as 1.

@ Data not available/ Numbers outside range of program

Table 3.7

GENE1C/2C Residual Analysis

Criterion	Expected	Computed	Residue	%Error
w/o Cat.Rep.	120.00026	61.19817	58.80209	49.0
0.4% ara	24.5	24.59973	.09973	.41
0.1 cAMP	24.00005	58.96817	34.96812	145.7
X1:X2 (I)	33.3:1	60:1 *	26.7 *	80.2
0/2x cAMP	134.18600	61.14675	73.03925	54.3
X2=0 (u)	1.0	4.24521	3.24521	324.5
10% S.T. (I)	300 sec.	34 sec.	266 sec.	
X4/X5 Diff.	0	10.9165	10.9165	
#-/0 values	0	6	6	

* X2 negative, taken as 1.

Table 3.8

GENE3A/3B Residual Analysis

Criterion	Expected	Computed	Residue	%Error
w/o Cat.Rep.	120.04806	199.7 @	79.65194	66.35
0.4% ara	24.5	27.53868	3.03868	12.40
0.1 cAMP	24.00961	54.72382	30.71421	127.92
X1:X2 (I)	33.3:1	60:1 *	26.7 *	80.18
0/2x cAMP	78.43054	34.96531	43.46523	55.42
X2=0 (u)	1.0	9.43011	8.43011	843.01
10% S.T. (I)	300 sec.	12 sec.	288 sec.	
X4/X5 Diff.	0	16.67338	16.67338	
#-/0 values	0	5	5	

* X2 zero, taken as 1.

@ Averaged over last period

Table 3.9

Model Discrimination

Residual Analysis

Criterion:	GeneOA/OB	Gene1/2	Gene1C/2C	Gene3A/3B
w/o cat. rep.	55.8524	38.85839	58.80209	79.65194
0.4% arabinose	.26729	.07117	.09973	3.03868
0.1 cAMP input	34.972	18.97391	34.96812	30.71421
X1:X2 I ratio	12.6	26.7*	26.7*	26.7*
0/2x cAMP input	73.2979	14.73957	73.03925	43.46523
X2=0 (u)	1.15432	@	3.24521	8.43011
10% settling (I)	295	56	266	288
X4/X5 differ	.00475	23.42576	10.9165	16.671421
# of -/0 values	3	6	6	5

* X2 value zero or negative

@ Data not available

Chapter 4

System Analysis

4.1 Introduction

The models developed in this study form linear time invariant representations of the L-arabinose system. Several techniques are available to study various aspects of the modeled system. These techniques include eigenvalue analysis, controllability analysis, frequency analysis and state space analysis. An eigenvalue analysis can be used to study the stability and transient behavior of the system. Formation of the controllability matrix can be used to test the controllability of the system. The frequency response of the system can be examined by deriving the transfer matrix of the system. Lastly, various state dependencies can be examined by appropriate state space plots.

Also discussed in this chapter is a further examination of GENEOA/OB and an attempt that was made to improve its performance.

4.2 Eigenvalue Analysis

Eigenvalues were computed, using equation 2.20, for the four models developed and for the original system (Figure 2.2). These eigenvalues are listed in Table 4.1. All models have two real eigenvalues and three pairs of complex eigenvalues. The real parts of

the eigenvalues of GENEOA/OB ranged from -1.373 to -.11, while those of GENE1/2 and GENE1C/2C ranged from -1.0 to -.01 and from -1.649 to -.072 respectively. In all models, except GENE3A/3B, all of the eigenvalues are located in the left half of the complex plane. For an LTI system, this fact is enough to establish the asmyptotic stability of the system [3].

Examination of the eigenvalues also shows the reason for the faster settling time and reduced oscillatory behavior of GENEOA/OB compared to the other models. All other models have eigenvalues closer to the imaginary axis than those of GENEOA/OB. The closer the eigenvalues are to the imaginary axis the longer lasting and more oscillatory the transient behavior becomes.

As noted in section 3.2, the oscillatory behavior of GENE3A/3B was not detected during model development. Further analysis of GENE3A/3B revealed that the uninduced system retained its oscillations when run for 1200 seconds with the output (y) ranging between zero and approximately eighteen. The apparent stability of the induced system, GENE3B, turned out to be due to the state limiter which restricted the states of the model to non negative values. This limiter was a program loop that reset any negative state values to zero at each iteration of the differential equation solver. Thus, this limiter introduced a nonlinearity into the system. When this introduced nonlinearity was removed, the instability of the system, predicted by the positive eigenvalues, became evident.

4.3 Controllability

A computer program was used to test the original system (Figure 2.2) for controllability. The program tested for controllability by forming the controllability matrix of the system. For an LTI system, the controllability matrix is defined as follows.

$$C = [B \mid AB \mid A^2B \mid \_\_\_ \mid A^{n-1}B] \qquad (4.1)$$

If the rank of the controllability matrix is n (the dimension of the system), the system is completely controllable. Stating that a system is controllable means that the states of the system can be moved from any arbitrary initial condition to any desired final values by use of the inputs to the system only [3].

For the original system, the rank of the controllability matrix (Figure 4.1) is n with the first eight columns of the matrix being linearly independant. Thus, the system as defined is indeed controllable, and it can be stated that the choice of inputs for the model are adequate to control the input/output behavior of the model. This is physiologically realistic since the three inputs chosen represent the major external substances involved with the control of the activity of the gene. Knowledge of the controllability also indicates that the selection of where to apply the inputs is sufficient for system control (ie. the non zero entries in the B matrix), although it does not rule out the possibility of the inputs having a direct effect on other state variables.

4.4 Frequency Analysis

One approach that can be used to analyze the behavior of an LTI system is to examine the frequency response of the system. This involves formulating the transfer function of the system and examining the response of the output to a sinusoidal input. This can be most easily done using the Laplace Transform of the system as follows.

For this LTI system with zero initial conditions [3].

$$X = AX(t) + BU(t) \qquad (4.2)$$

$$y = CX(t) \qquad (4.3)$$

the Laplace transformation yields:

$$sX(s) = AX(s) + BU(s) \qquad (4.4)$$

$$Y(s) = CX(s) \qquad (4.5)$$

Solving for the input/output relationship yields:

$$(sI-A)X(s) = BU(s)$$

$$X(s) = (sI-A)^{-1}BU(s)$$

$$Y(s) = C(sI-A)^{-1}BU(s)$$

$$H(s) = Y(s)/U(s) = C(sI-A)^{-1}B \qquad (4.6)$$

$$H(s) = N(s)/D(s) \qquad (4.7)$$

H(s) is called the system transfer matrix. For a single input single output (SISO) system, H(s) reduces to the system transfer function ,n(s)/d(s), where d(s) represents the characteristic equation of the system. The roots of d(s)=0 represent the eigenvalues (poles) of the system, and the roots of n(s)=0 represent the zeros of the system.

For GENEOA/OB, where there are three inputs, H(s) can be written as follows.

$$H(s) = (1/\det(sI-A))\ P(s) \qquad (4.8)$$

In this equation, P(s) is a 1x3 matrix whose elements are polynomials in s, and the det(sI-A) represents the common denominator of the elements of H(s). For GENEOA/OB, the det(sI-A) is an eighth order polynomial that represents the characteristic equation for the system, and the roots of det(sI-A)=0 represent the eigenvalues of the system. Each column of P(s) is a polynomial in s of order less than eight that represents the numerator of the transfer function relating one of the inputs to the system output. H(s) for GENEOA/OB is shown in Figure 4.2.

The steady state frequency response of the system can now be obtained by letting s=jw in H(s). Thus forming the frequency response matrix of the system H(jw). Since our system is asymptotically stable, the response of the system to a constant input (U_p) is given by:

$$y(t) = H(0)U_p$$

For GENEOB, if we let jw=0 in H(jw) and let our inputs equal 1.0 (our original assumption), we find that y(t)=57.798 which differs from the computer generated steady state response of the system only due to round off error.

Multivariate frequency domain analysis techniques that correspond to those for SISO systems (Nyquist and Bode Plots) can be used for studying the frequency response of this three input model. However, those techniques, such as the Multivariable Nyquist Criteria, are

relatively new and were not used in this study [34].

4.5 State Space Analysis

The analysis of GENEOA/OB continued with an examination of selected state dependencies. The activity of PBAD, the promoter for the B, A and D genes, in relation to the activity of Pc, the promoter for the C gene, was investigated for both the uninduced and induced system (Figures 4.3 and 4.4). From the plots of X_1 vs X_2, it can be seen that in the uninduced system, the two states track each other with an apparent phase lag to their steady state values, while in the induced system, their relationship is not well defined.

Figures 4.5 and 4.6 show the time dependencies in the uninduced system of the free and bound araC protein respectively. The free araC protein represents the sum of the concentrations of the repressor and activator ($X_7 + X_8$). The bound protein is the sum of the associations of the protein to araO and araI ($X_3 + X_4 + X_5$). Figures 4.7 and 4.8 show the same data for the induced system. It can be seen by comparing these plots to those plots for the corresponding output of the system (Figures 4.9 and 4.10), that the "protein states" follow the same trajectory as the overall response of the system.

When the repressor (P1) is plotted against the activator (P2) for the uninduced and induced system, Figures 4.11 and 4.12 respectively, the results are very similar to those plots of PBAD vs Pc activity. That is, in the uninduced system, the repressor and activator

concentrations track each other with an apparent phase lag toward their steady state values. While in the induced system, their balanced reaction is disturbed. Again, similar results are obtained when the output is plotted against the free protein (Figures 4.13 and 4.14).

Plotting the output vs araI activity (Figures 4.15 and 4.16) reveals the tracking in the uninduced system once more, but also shows a higher degree of tracking in the induced system than seen in other state dependency plots.

4.6 GENEOA/OB Analysis

The model, GENEOA/OB, was selected in Chapter 3 as the model for the cost analysis portion of this research. Therefore, further analytical investigation was restricted to this model. The model (Figure 3.1) was transformed into the Jordan Normal form, as shown in Figure 4.17, and an eigenvalue analysis was performed as in section 2.2. Through the eigenvalue sensitivity analysis (Table 4.2), it was revealed that the uninduced system response was most sensitive to changes in the system's real eigenvalues as fifty percent changes in the eigenvalues caused the largest percent changes in the output (y) especially for the uninduced system. It also revealed that the induced system response was relatively insensitive to changes in the eigenvalues. Using the real eigenvalues, the system was optimized as shown in Table 4.3. In this table, as in Chapter 2, e represents the system error. This model was then returned to the original state space representation by the inverse transform. The resulting new A matrix

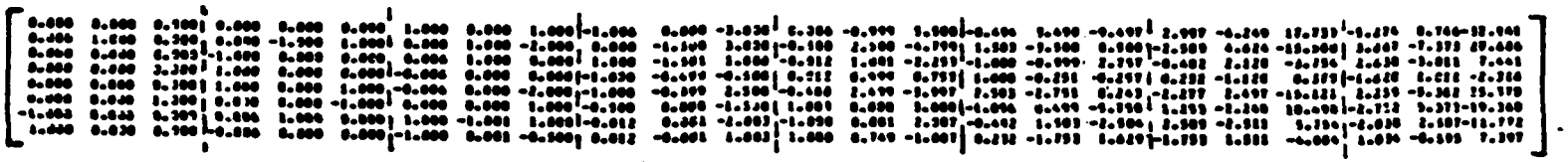
for the model is shown in Figure 4.18. Here again, as in GENE1/2, the state dependencies have changed dramatically due to the eigenvalue optimization.

This new model, GENE7A/7B, was evaluated, as in Chapter 3, against the same criteria as the original four models with the results presented in Table 4.4. GENE7A/7B was a better fit model than GENEOA/OB for 6 of the 9 criteria. Although, comparisons between GENEOA/OB and GENE7A/7B in table 4.4 show mostly minor differences, there are significant differences for the X_1:X_2 ratio and the response when X_2 equals zero. The fact that most differences were minor and the physiological anomaly of GENE7A/7B (ie. the lack of X_3 dependency upon X_7) suggested that GENEOA/OB was still the better choice for the system model. The smaller residual in the X_4/X_5 difference and the faster settling time of GENEOA/OB also indicate that GENEOA/OB is still the best choice for the system model.

One area, in which all models showed significant deviation from the reported data, was the response of the system to the absence of catabolite repression. When the models were modified to remove the dependency of cAMP binding upon the catabolism of L-arabinose (X_6 no longer dependant upon X_1), the increase in X_1 activity and hence, enzyme production was significantly below that predicted in the literature [31]. In the biological system, the removal of catabolite repression involved the removal of the catabolites (the products of the action of the BAD enzymes) that were reducing the intracellular

concentrations of cAMP. In the model, catabolite repression was simulated by the inverse dependency of X_6 upon X_1, so that removal of catabolite repression involved deleting the dependency of X_6 upon X_1 by letting coefficient a_{17} in equation 2.8 equal zero.

An attempt was made with GENEOA/OB to improve this response by increasing the dependancy of X_6 upon X_1 and simultaneously increasing the cAMP input. The results, as shown in Table 4.5, indicate little improvement of the model using this approach. Although, further adjustments using this technique might be of use in the model, it is felt that a totally different approach might be necessary. This could possibly involve the addition of another state which reflects cAMP concentration alone. One could then relate the cAMP(CAP)/DNA associations to this new state variable.



Controllability Matrix of Original Model

Figure 4.1

$$A(s)=311.73s^5 + 948.9s^4 + 1324.9s^3 + 1329.8s^2 + 849.09s + 207.0$$

$$B(s)= -23.468s^3 -35.252s^2 -10.63s + 1.154$$

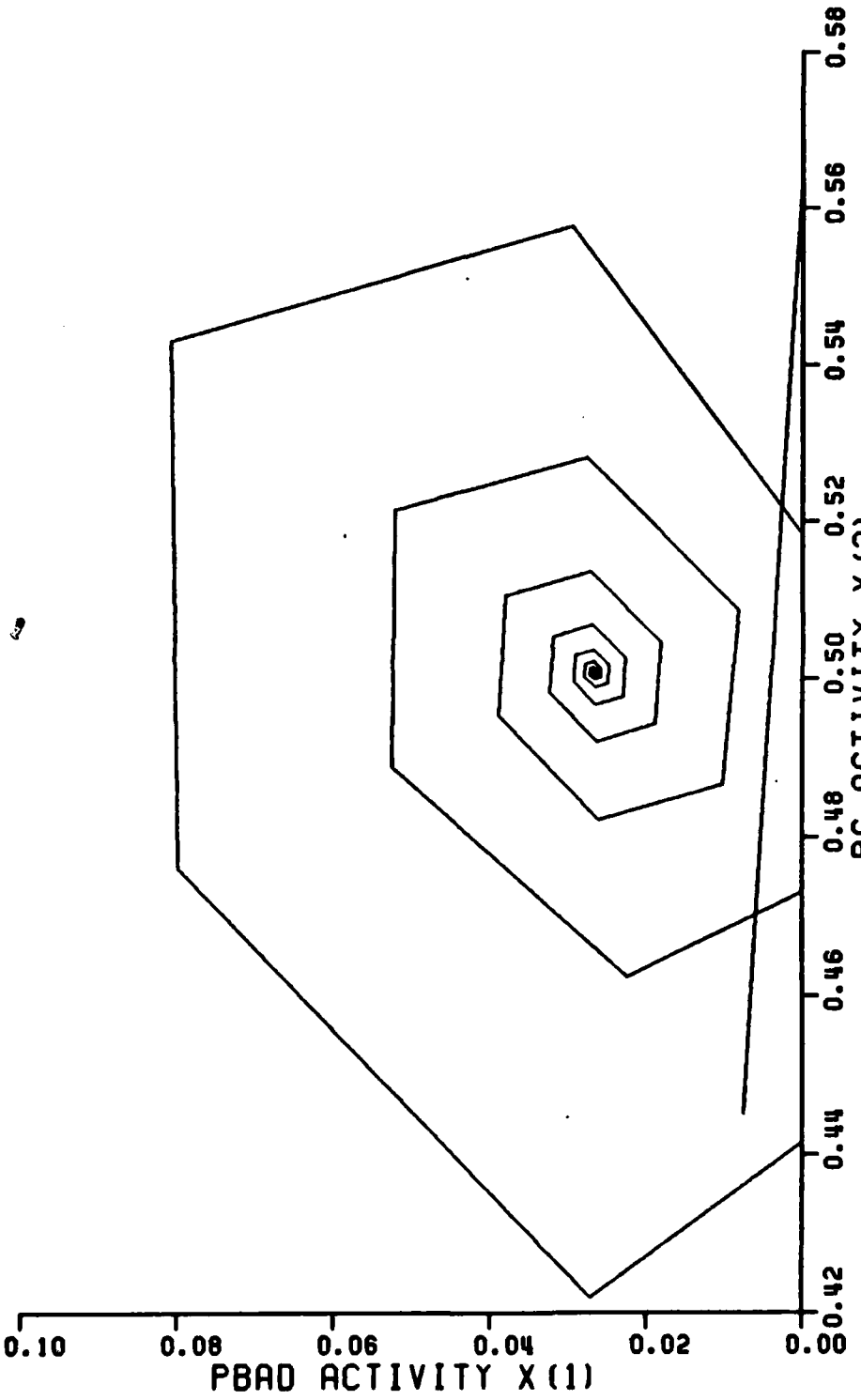
$$C(s)=5.3952s^5 + 10.823s^4 + 17.06s^3 + 18.684s^2 + 11.316s + 9.1665$$

$$D(s)=s^8 + 5.006s^7 + 14.78s^6 + 30.793s^5 + 43.736s^4 + 44.815s^3 + 33.982s^2 + 16.462s + 3.760$$

$$H(s) = \frac{\left[A(s) \quad B(s) \quad C(s)\right]}{D(s)}$$

Transfer Matrix, H(s), for GENEOA/OB

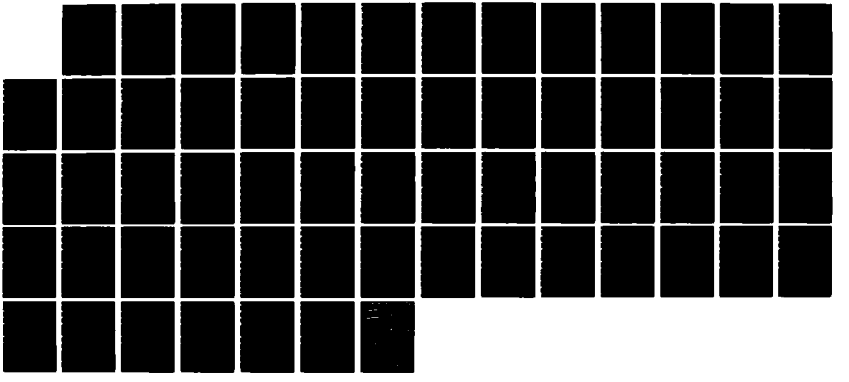
Figure 4.2

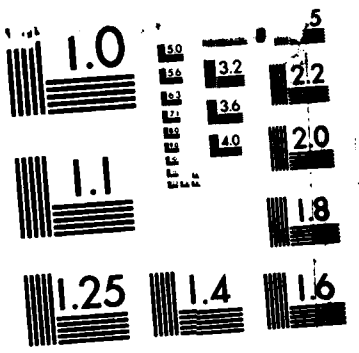


PBAD vs Pc for GENEOA
Figure 4.3

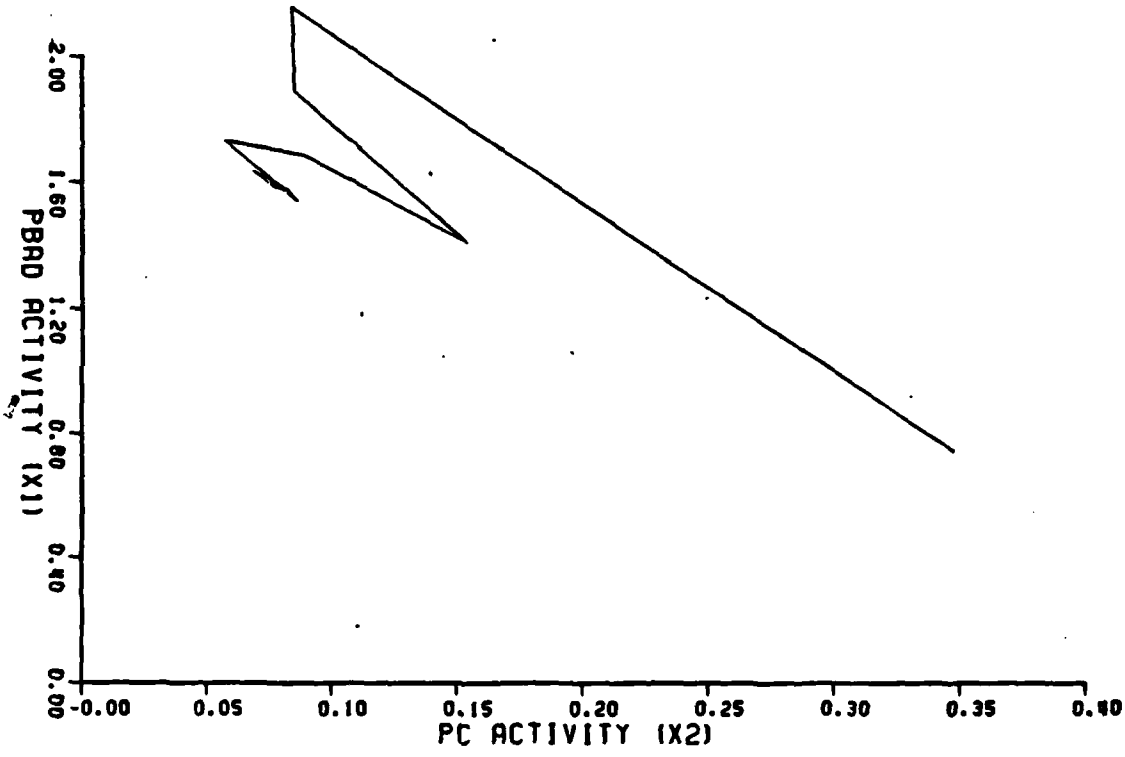
AD-A166 374 A COMPUTER SIMULATION OF THE L-ARABINOSE GENE-ENZYME COMPLEX WITH AN ANAL (U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH B L GEORGE 1985 2/2

UNCLASSIFIED AFIT/CI/NR-86-6D F/G 6/1 NL

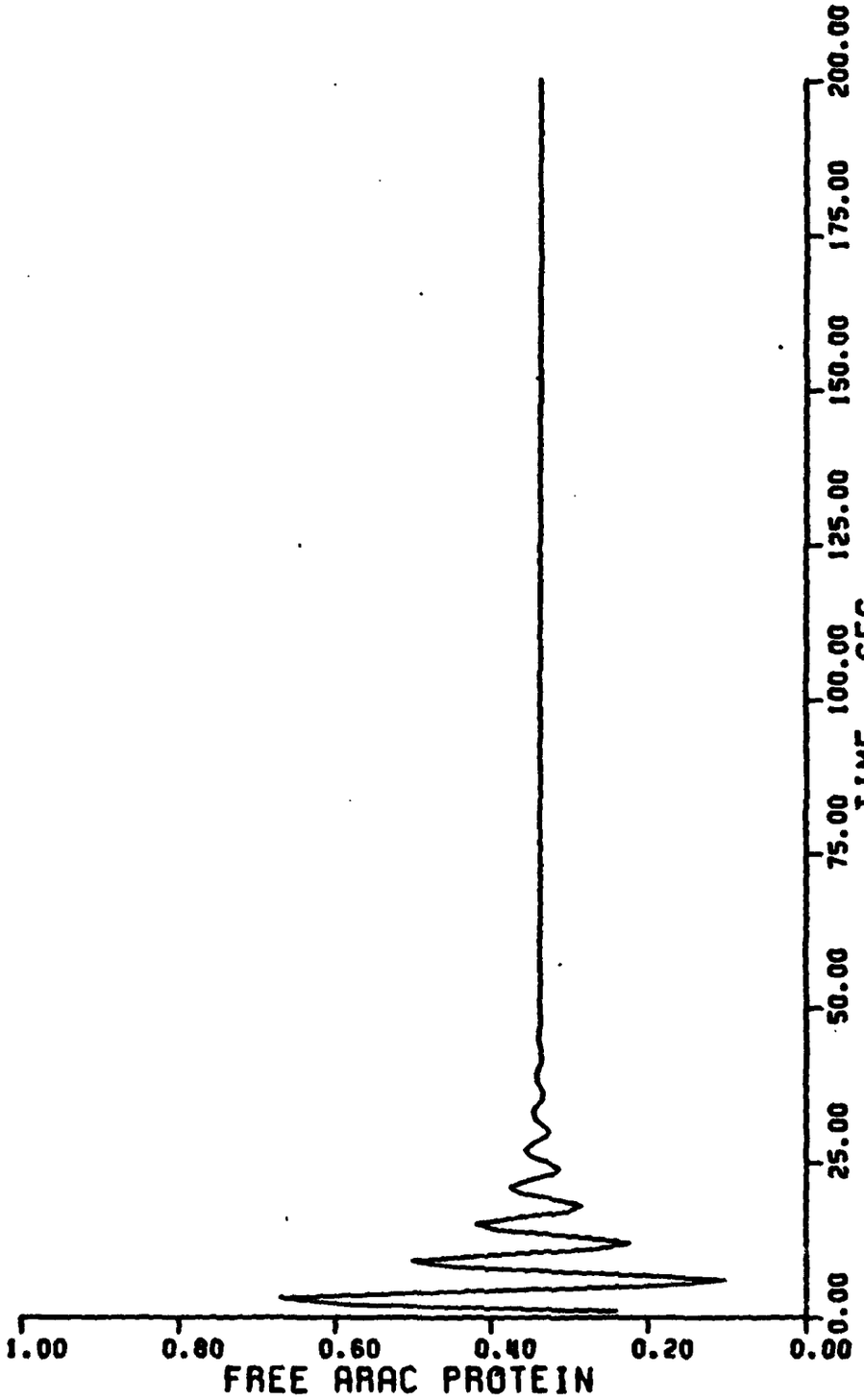




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

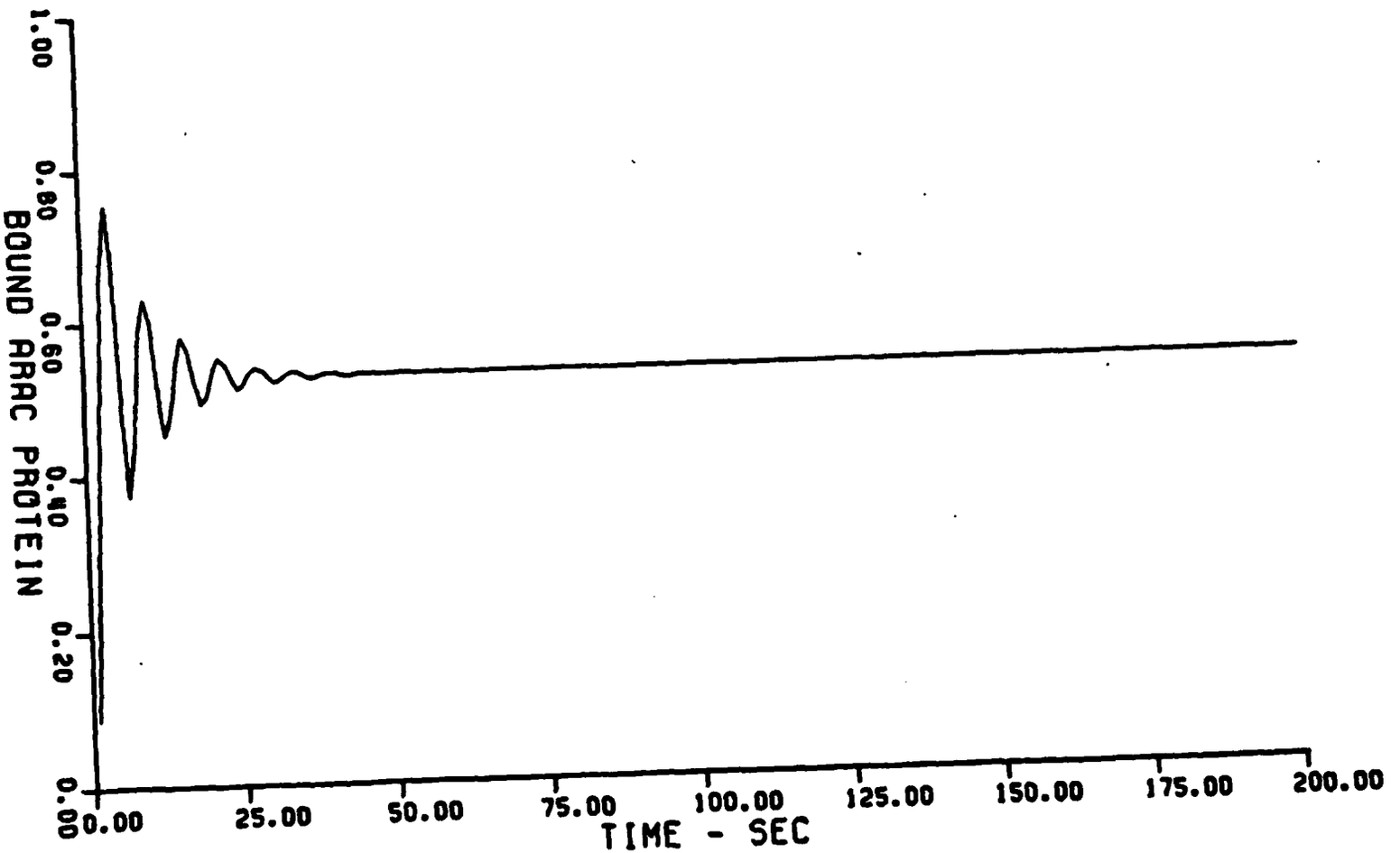


PBAD vs Pc for GENEOB
Figure 4.4



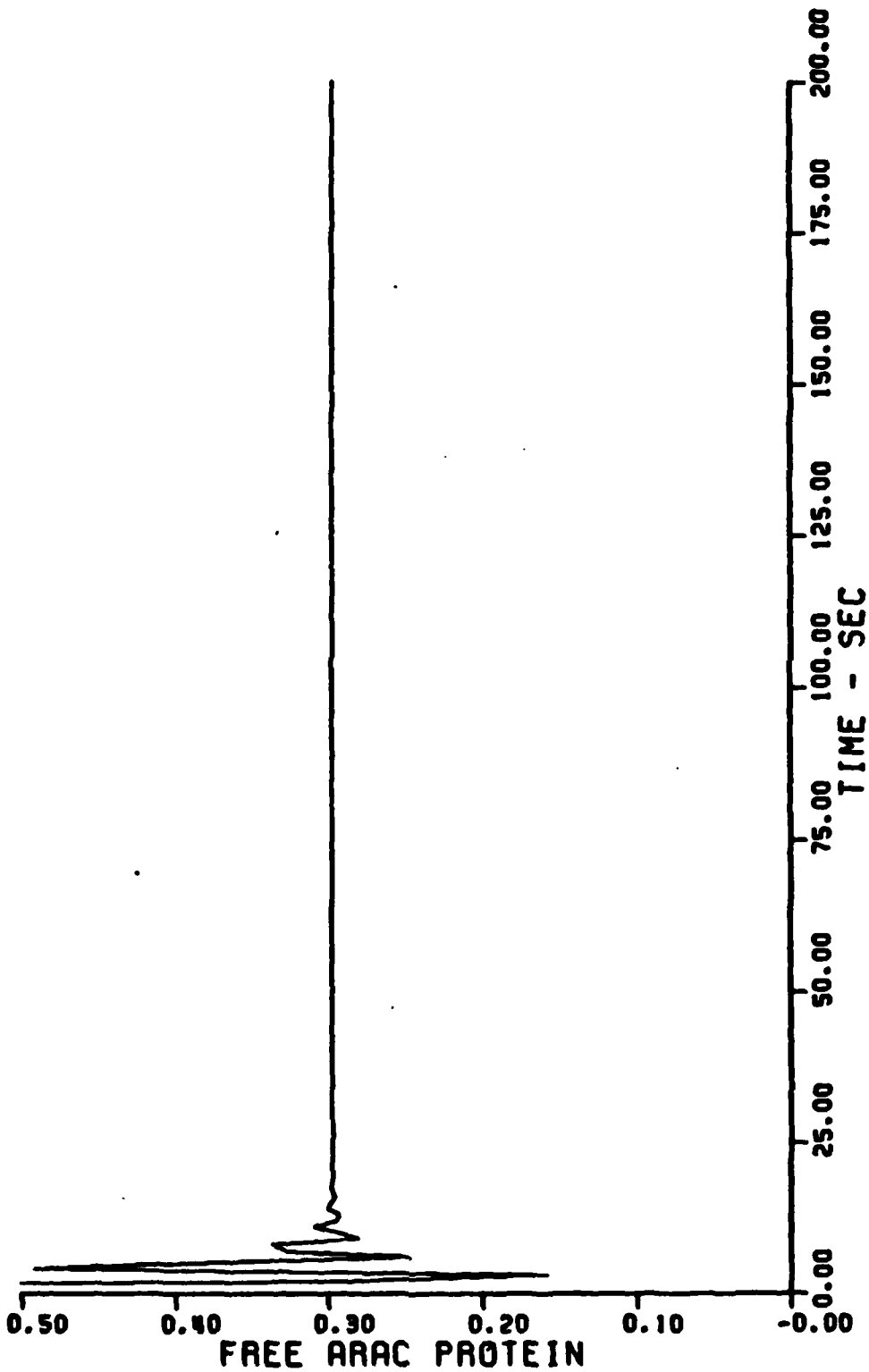
Free Protein vs Time for GENEOA

Figure 4.5



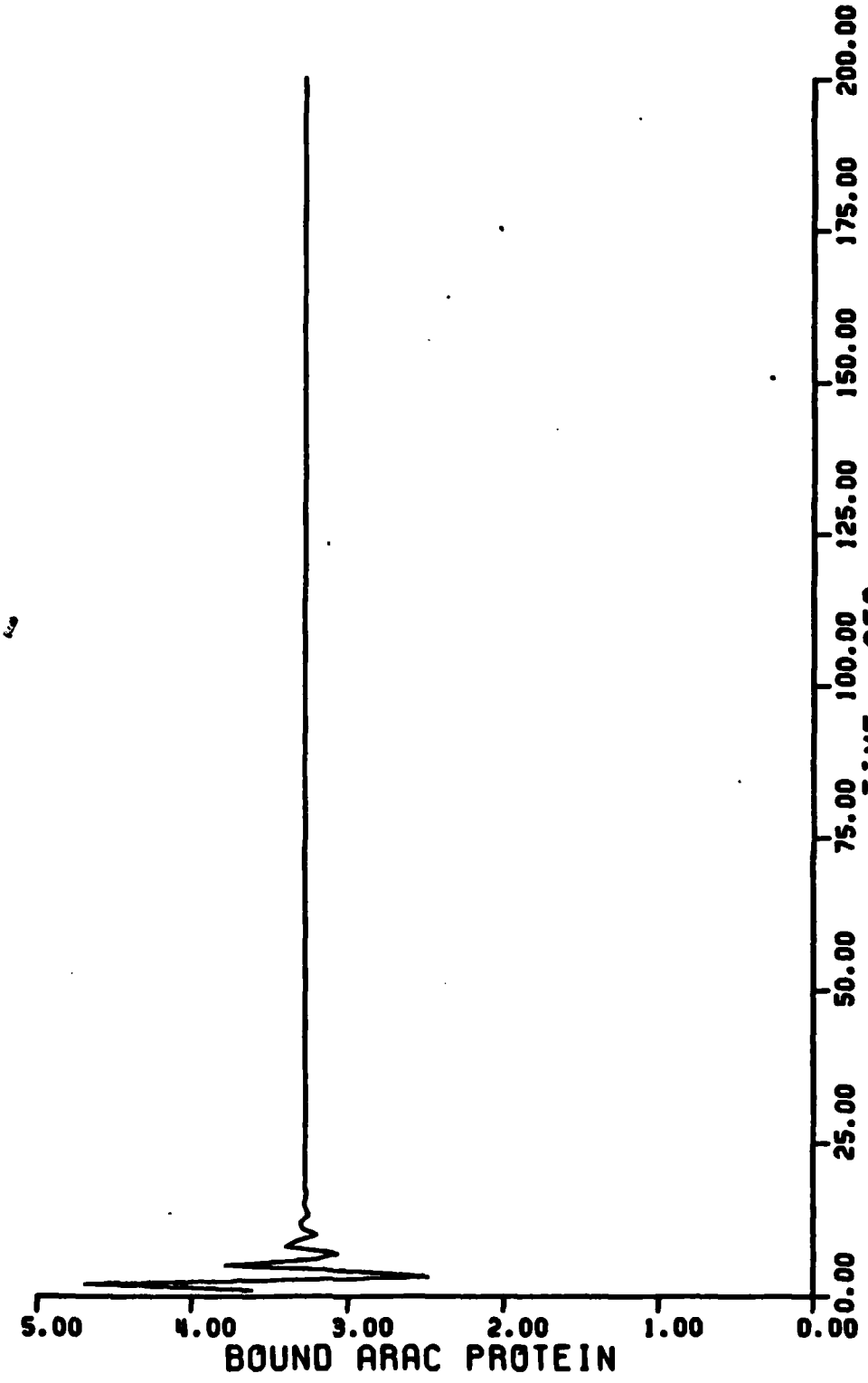
Bound Protein vs Time for GENEOA

Figure 4.6



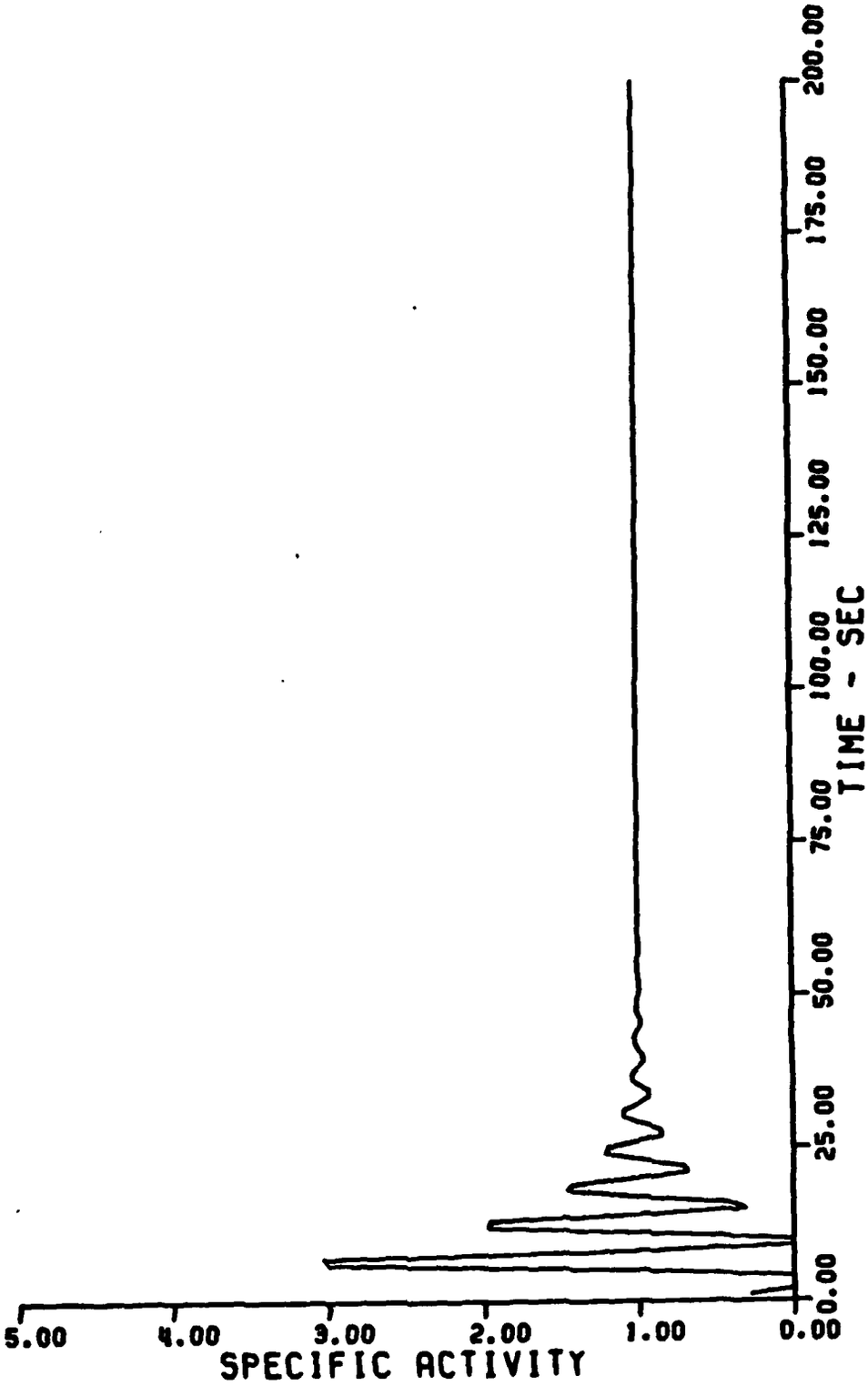
Free Protein vs Time for GENEOB

Figure 4.7



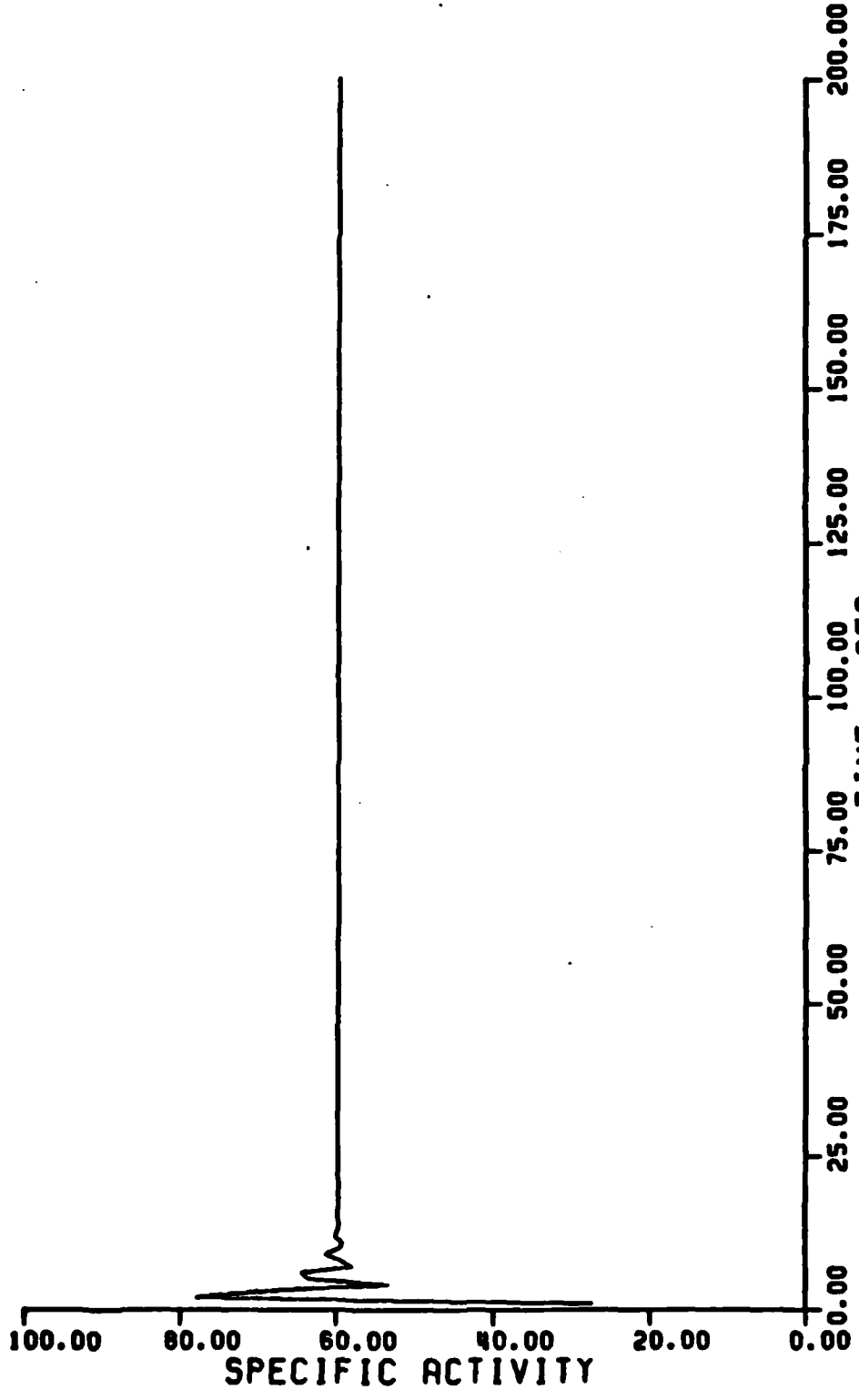
Bound Protein vs Time for GENEOB

Figure 4.8



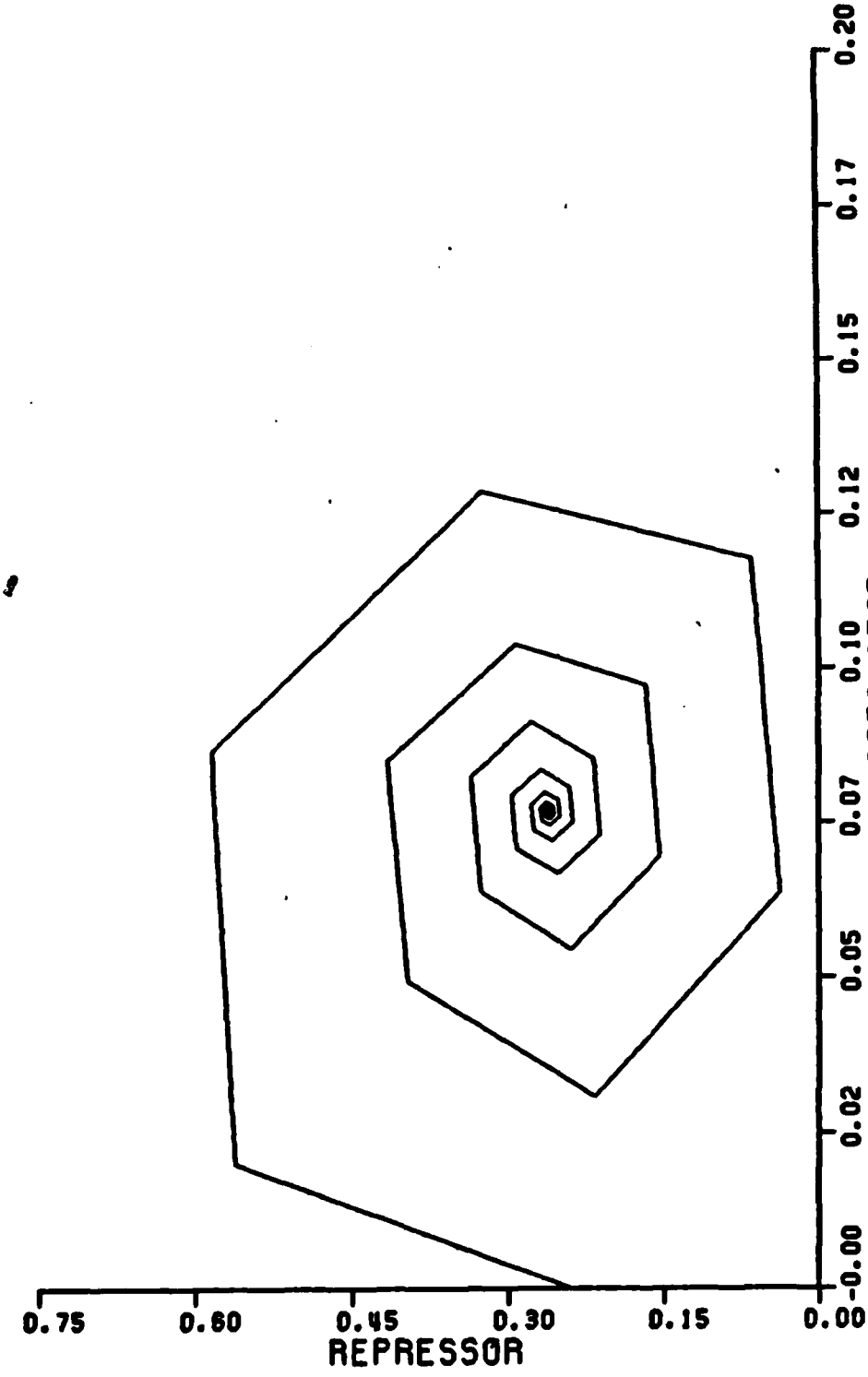
Uninduced Response of GENEOA

Figure 4.9



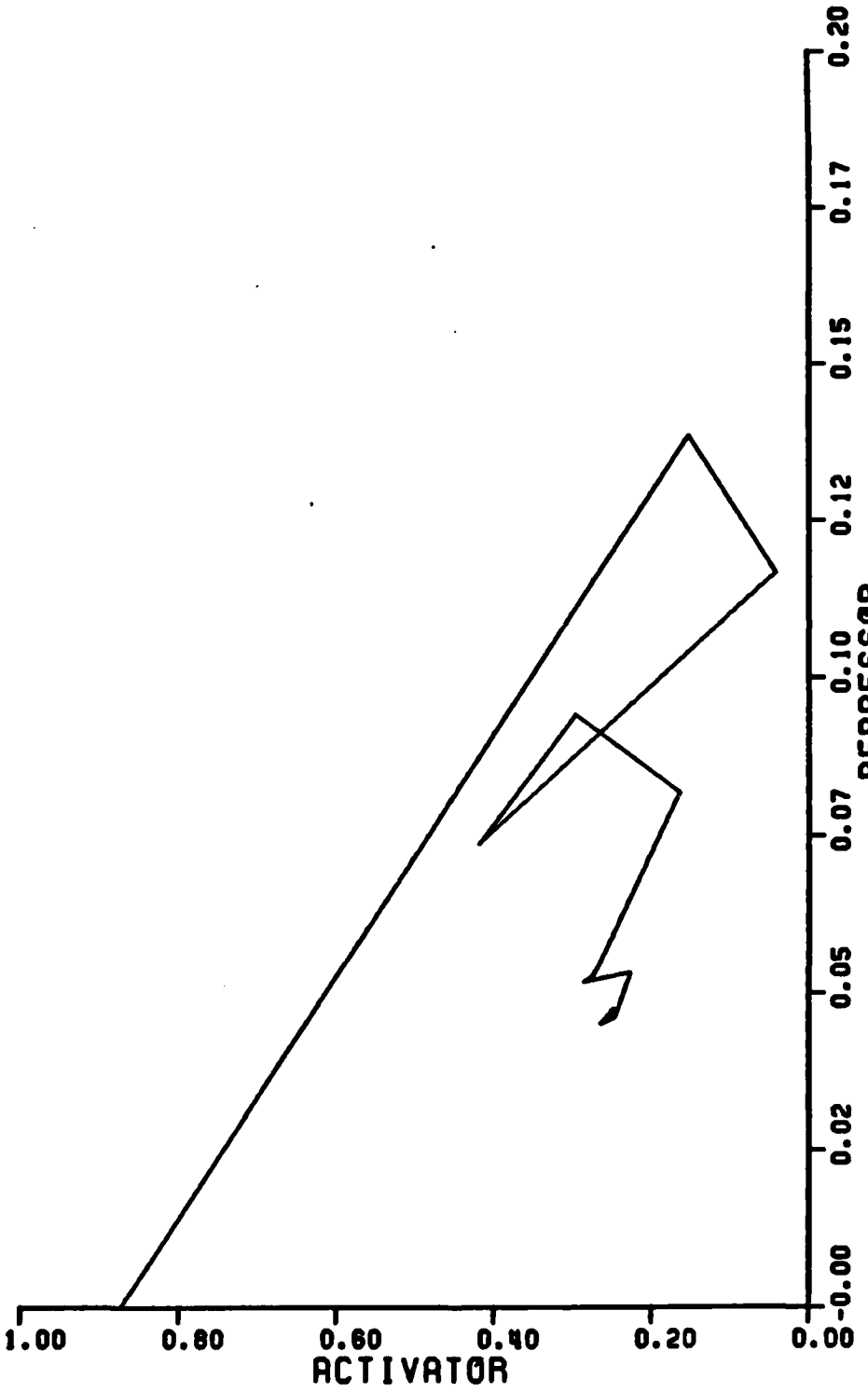
Induced Response of GENEOB

Figure 4.10



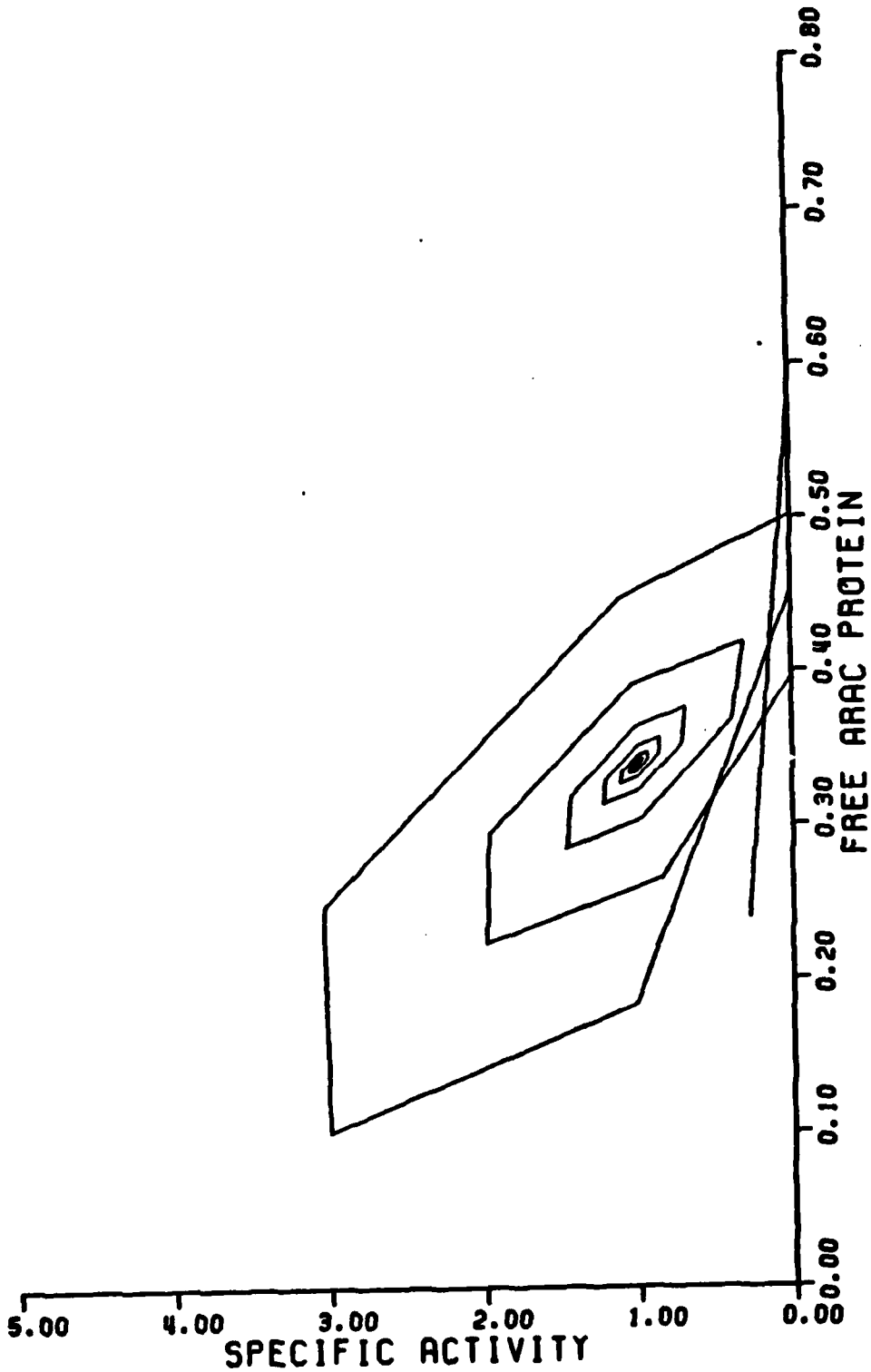
P1 vs P2 for GENEOA

Figure 4.11



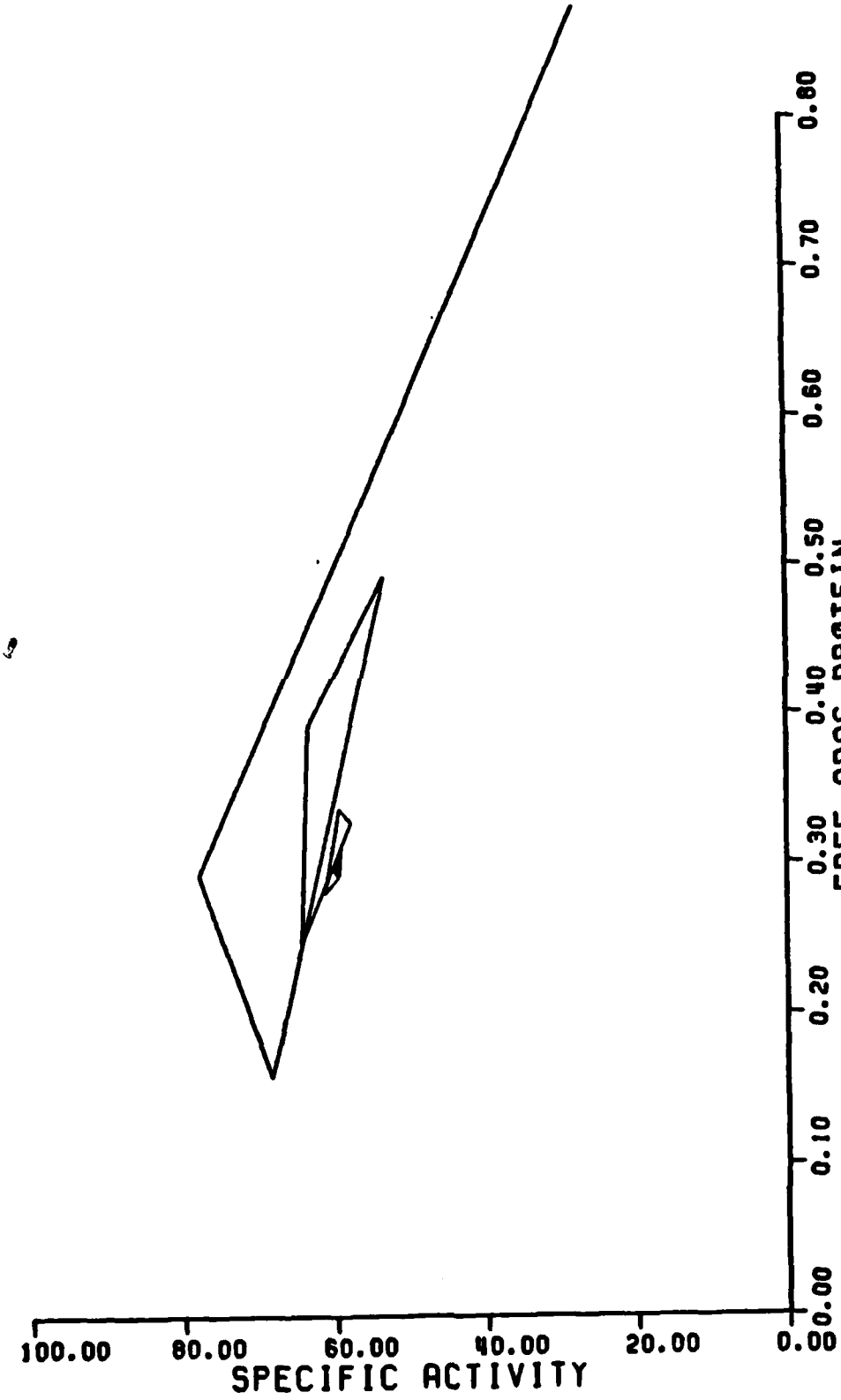
P1 vs P2 for GENEOB

Figure 4.12



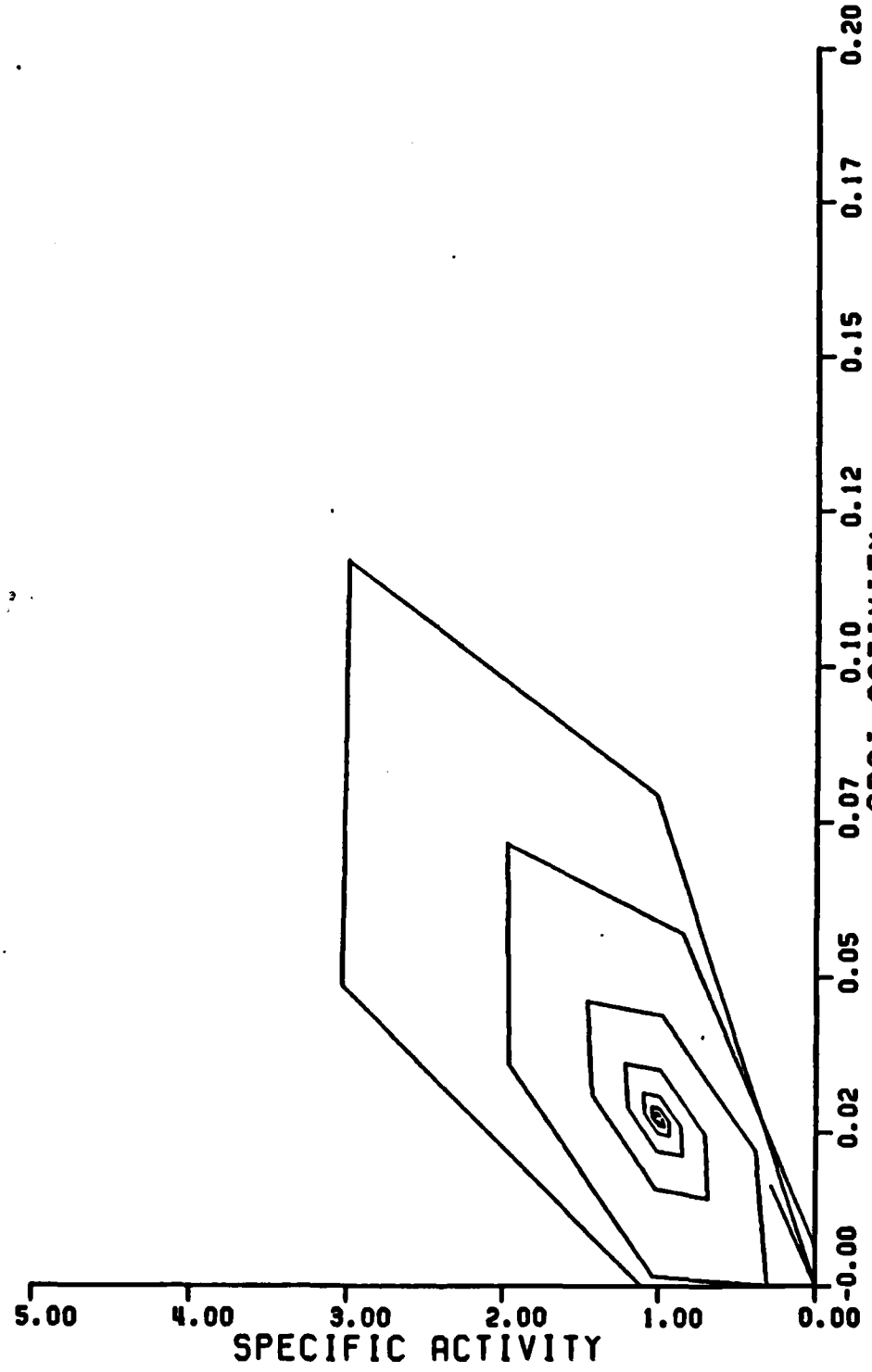
Output vs Free Protein for GENEOA

Figure 4.13



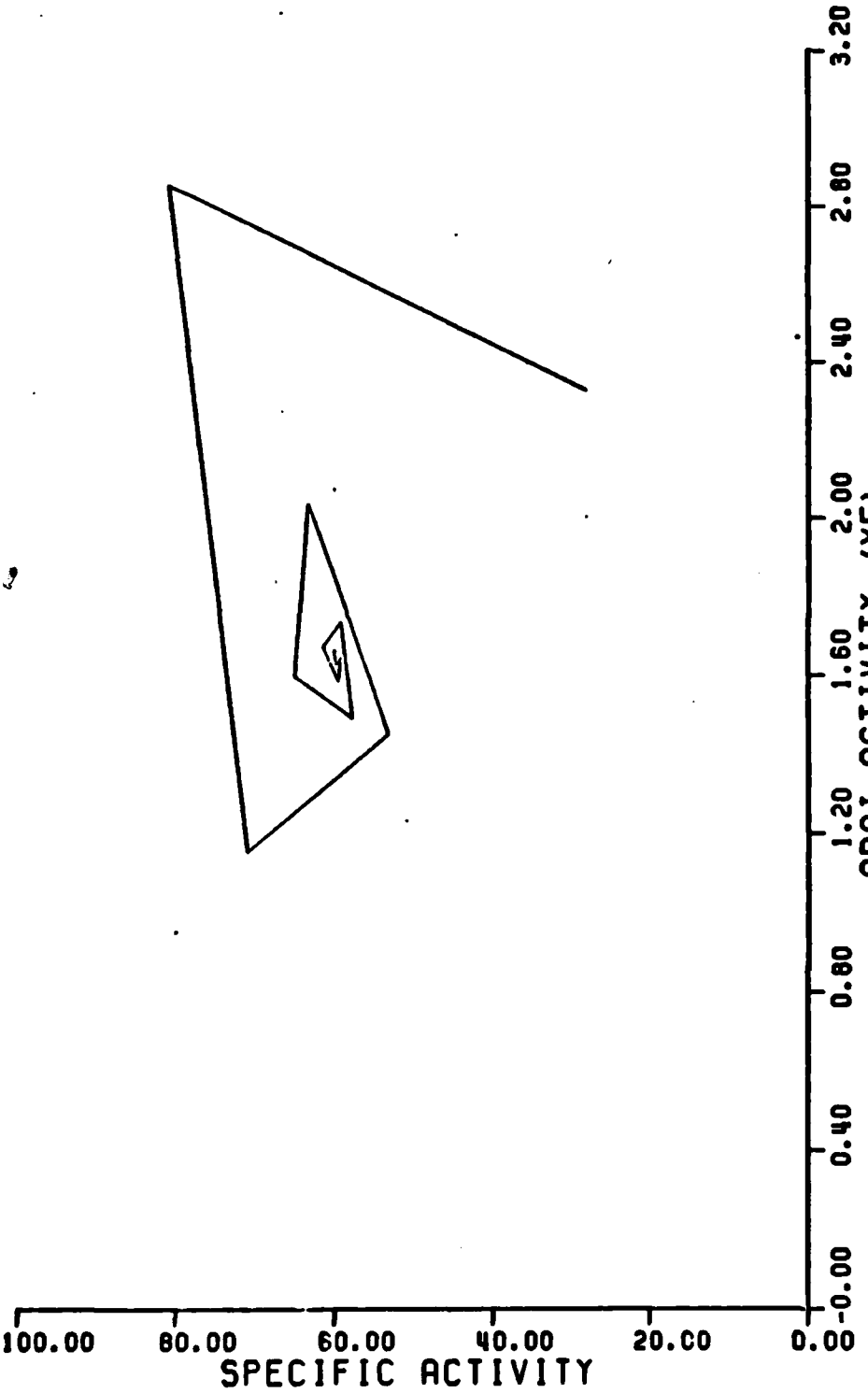
Output vs Free Protein for GENEOB

Figure 4.14



Output vs araI for GENEOA

Figure 4.15



Output vs arai for GENEOB

Figure 4.16

$$\begin{bmatrix}
-0.411 & 1.850 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
-1.850 & -0.411 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -0.110 & 1.049 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -1.049 & -0.110 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.649 & 0.329 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.329 & -0.649 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.373 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.294
\end{bmatrix}$$

Real Jordan Normal Form of GENEOA/OB

Figure 4.17

$$\begin{bmatrix}
-0.933 & -0.142 & -0.070 & -0.052 & 0.992 & 0.048 & 0.047 & 0.128 \\
0.024 & -1.113 & -0.549 & -0.565 & 0.017 & 0.115 & 0.034 & 0.058 \\
0.017 & -0.067 & -0.329 & -0.537 & 0.008 & 0.011 & 0.920 & 0.035 \\
0.017 & -0.044 & -0.521 & -0.520 & 0.000 & 0.012 & 0.014 & 2.534 \\
-0.024 & 0.047 & -0.977 & 0.016 & -0.997 & 0.382 & -0.016 & 5.156 \\
-0.404 & -0.220 & -0.106 & -0.099 & -0.005 & -0.933 & 0.071 & 0.184 \\
-0.006 & 1.035 & -0.985 & 0.022 & -0.007 & -0.003 & -0.011 & -0.011 \\
-0.002 & 0.001 & 0.001 & -0.501 & -0.499 & -0.002 & 0.000 & -0.009
\end{bmatrix}$$

New A Matrix for GENE7A/7B

Figure 4.18

Table 4.1

Eigenvalues

Original	GENE0A/0B	GENE1/2	GENE1C/2C	GENE3A/3B
-1.809	-.411+1.85j	-.045	-.072+.295j	-1.573
-.067+1.067j	-1.373	-.01+.108j	-1.004	-.436+.603j
-.701+.601j	-.11+1.049j	-1.0	-.557+.136j	-.582
-.331+.861j	-1.294	-.044+.019j	-1.649	.229+1.153j
-1.0	-.649+.329j	-.701+.45j	-.302+.409j	.072+.232j

Table 4.2

GENE7A/7B Analysis

Nominal Values:

Y_U = 2.70252 Y_I = 57.79885

Eigenvalue Sensitivity:

Eigenvalue	%Change	Y_U	%Y	Y_I	%Y
EG1(R)-.370	-50	2.74069	1.4	58.81736	1.8
-.452	+50	2.66430	1.4	56.79319	1.7
EG1(I) 1.665	-50	2.74578	1.6	57.28474	.9
2.035	+50	2.66005	1.6	58.04540	.4
EG2 -1.236	-50	1.36707	49.4	60.98005	5.5
-1.512	+50	3.07216	13.7	55.19507	4.5
EG3(R)-.099	-50	2.69041	.4	57.72588	.1
-.121	+50	2.71551	.5	57.87167	.1
EG3(I) .944	-50	2.23844	17.2	57.80063	0
1.154	+50	3.08528	14.2	57.81067	0
EG4 -1.165	-50	3.53919	31.0	57.82901	.1
-1.423	+50	1.29406	52.1	57.77416	0
EG5(R)-.584	-50	3.40098	25.8	56.66299	2.0
-.714	+50	2.14669	20.6	58.30939	.9
EG5(I) .296	-50	2.03328	24.8	60.96844	5.5
.362	+50	2.54582	5.8	54.78762	5.2

Table 4.3

GENE7A/7B

Eigenvalue Optimization:

EG2	Y_I	e_I	Y_U	e_U
-1.274	60.02922	.00085		
-1.275	60.00497	.00002	2.05231	1.10736
-1.276	59.98076	.00037		
EG4				
-1.428			1.00617	.00004
-1.429	59.97913	.00044	.99911	.00000
-1.430			.99205	.00006

Table 4.4

GENEOA/OB & GENE7A/7B

Residual Analysis

Criteria	GENEOA/OB	GENE7A/7B
w/o cat. rep.	55.8524 (46.75%)	54.61397 (45.43%)
0.4% arabinose	.26729 (1.09%)	.15681 (.64%)
0.1cAMP input	34.972 (146.4%)	33.54954 (139.5%)
X_1:X_2 ratio	12.6 (37.84%)	33.63 (101.0%)
0/2x cAMP	73.2979 (54.70%)	67.7666 (51.86%)
X_2=0	1.15432 (115.4%)	.15334 (15.33%)
10% settling (U)	215	295
X_4/X_5 difference	.00475	.14555
#-/0 values I/U	3	3

Residue = |Expected value - Computed value|

Percentages = (residue/expected value) x 100

Table 4.5

GENEOA/08
Catabolite Repression Analysis

Criteria	a_{17} = -.75	a_{17} = -.50
Normal cAMP	Y_I=59.99834 e_I=.0000	Y_I=59.73076 e_I=.07249
0.1 cAMP input	Expected=23.99934 Computed=57.96500 Residual =33.9657 (141.5%)	Expected=23.8923 Computed=58.86434 Residual =34.972 (146.4%)
w/o cat. rep.	Expected=119.9967 Computed=67.31602 Residual =52.6807 (43.9%)	Expected=119.4615 Computed=63.60912 Residual =55.8524 (46.8%)
0/2x cAMP	Expected=132.1602 Computed=64.18052 Residual =67.9797 (51.44%)	Expected=133.9913 Computed=60.69341 Residual =73.2979 (54.7%)

Chapter 5

Repressor Model Development

In order to compare the cost of control, it was necessary to develop a hypothetical repressor only model for the L-arabinose system. This model would have to reproduce the input/output behavior of the system while using only repressor control. A five state model was selected with the following state assignments.

x_1 = PBAD/RNAP associations

x_2 = P_R/RNAP associations

x_3 = R/araO associations

x_4 = cAMP(CRP)/DNA associations

x_5 = Concentration of Repressor (R)

The first four states were selected because they represent the controlling sites necessary for a repressor controlled gene with cAMP/CRP activation.

P_R represents the promoter for the controlling gene, araR, (analogous to araC) which codes for the repressor (R). States x_3, x_4 and x_5 in this model are analogous to states x_3, x_6 and x_7 respectively in the activator/repressor model. Because there is no activator in this model, states x_4, x_5 and x_8 of the activator/repressor model are not represented in this model.

It is proposed that the operation of the system can be described by the following set of equations.

$$\dot{x}_1 = -a_1x_1 - a_2x_3 + a_3x_4 + b_1u_2 \qquad (5.1)$$

$$\dot{x}_2 = -a_4x_2 - a_5x_3 + a_6x_4 + b_2u_2 \qquad (5.2)$$

$$\dot{x}_3 = -a_7x_3 + a_8x_5 \qquad (5.3)$$

$$\dot{x}_4 = -a_9x_1 - a_{10}x_4 + b_3u_3 \qquad (5.4)$$

$$\dot{x}_5 = a_{11}x_2 - a_{12}x_5 - b_4u_1 \qquad (5.5)$$

In these equations, states are fed back upon themselves with coefficients a_1, a_4, a_7, a_{10} and a_{12}. The effect of the repressor upon PBAD and P_R (the gene which produces the repressor) is expressed with coefficients a_2 and a_5. Catabolite repression is simulated using coefficient a_9. Coefficient a_{11} provides repressor binding at araO. RNAP is input to both state x_1 and x_2 in this model; with no activator, PBAD would presumably be active simply in the absence of the repressor requiring only cAMP activation and a source of RNAP. The effect of L-arabinose in this model would be to deactivate the repressor substance. This is a classic approach incorporated in the original Jacob-Monod inducible operon, in which activity of the genetic system is stimulated when a substrate substance interacts and inactivates a repressor molecule. Using initial values as assumed in the eight state space model, this repressor model can be represented in state space form as follows.

$$\dot{X} = \begin{bmatrix} -1.0 & 0 & -1.0 & 1.0 & 0 \\ 0 & -1.0 & -1.0 & 1.0 & 0 \\ 0 & 0 & -1.0 & 0 & 1.0 \\ -.5 & 0 & 0 & -1.0 & 0 \\ 0 & 1.0 & 0 & 0 & -1.0 \end{bmatrix} X + \begin{bmatrix} 0 & 1.0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1.0 \\ -1. & 0 & 0 \end{bmatrix} U$$

$$y = [\ 37.4672\ 0\ 0\ 0\ 0\ ]X$$

Here again, as in GENEOA/OB, the output was scaled by c_1 in the output matrix. Upon implementation of this model, GENE4A/4B, on the computer, the following values were obtained.

	uninduced	Induced
x_1	.8	1.2
x_2	.8	1.2
x_3	.8	.2
x_4	.6	.4
x_5	.8	.2
Y	29.97375	44.96063

A sensitivity analysis and optimization were done first on the B matrix parameters (Table 5.1), which yielded the following system response.

$Y_u = 1.04908 \quad e_u = .00241$

$Y_I = 59.94751 \quad e_I = .00276$

Next, a sensitivity analysis was performed on the A matrix parameters. The analysis (Table 5.2) revealed that the uninduced system was very sensitive to small changes in the coefficients, while the induced system was relatively insensitive to these changes.

Because of these results and the already low values of system error, none of the A matrix parameters were changed. Characteristics of GENE4A/4B are presented in Table 5.3, and output plots are found in Figures 5.1 and 5.2.

GENE4A/4B was then tested, where applicable, against those extrinsic criteria used earlier with the results shown in Table 5.4. As presented, the results for GENE4A/4B compare favorably with those for GENE0A/0B except for the case when x_2 is set to zero.

It will be noted later that this model produced high costs compared to the activator/repressor model for the L-arabinose system (see chapter 6). To ensure that these high costs were not an artifact of simulation, a second repressor only model was developed (GENE8A/8B).

GENE8A/8B was developed from the same original state space model as GENE4A/4B, but with the final input values of GENE0A/0B as the initial B matrix. This resulting model produced the following response.

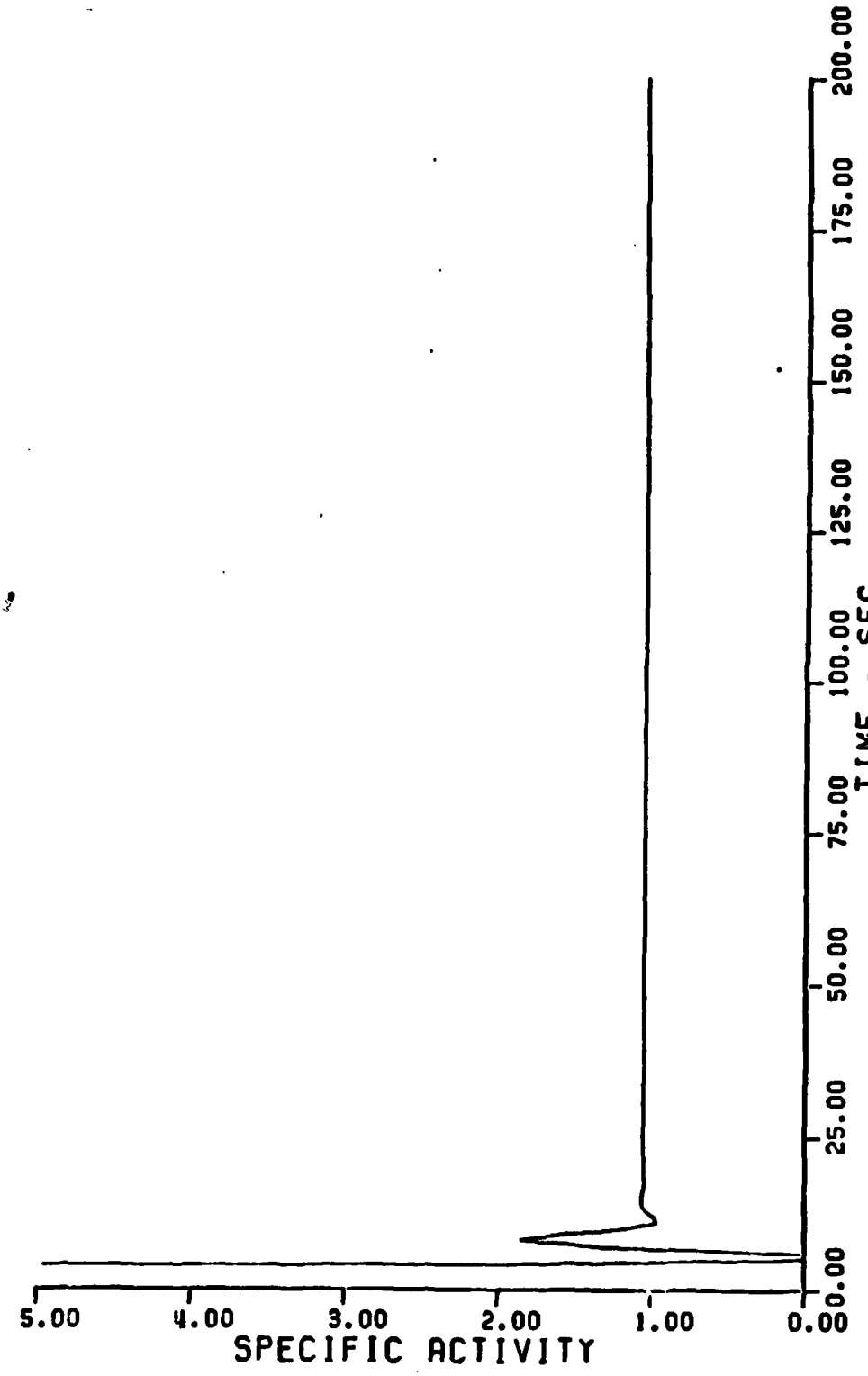
	uninduced	Induced
x_1	.424	.73337
x_2	.424	.73337
x_3	.424	0
x_4	.148	0
x_5	.424	0
Y	15.38609	27.47740

From this point, the model was optimized, first with the B matrix parameters and then with the A matrix parameters to yield the model in Figure 5.3 and with the following response (see also Figures 5.4 and 5.5).

$Y_u = 1.11345 \qquad e_u = .01403$

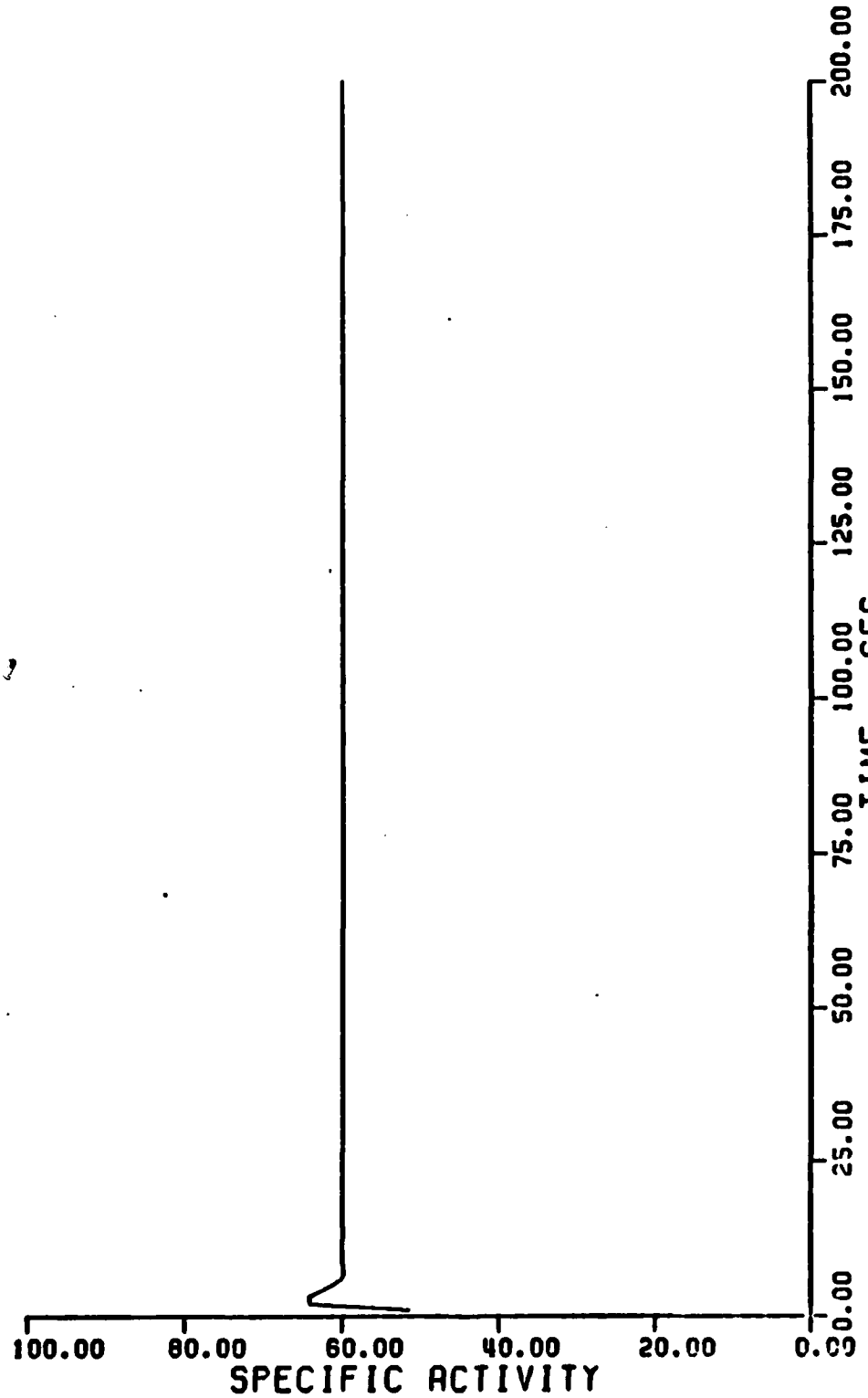
$Y_I = 60.17121 \qquad e_I = .02931$

A residue analysis done on GENE8A/8B produced the results in Table 5.6. Again, this repressor only model compared favorably with GENE0A/0B for all criteria except the case when x_2 equaled zero in the uninduced system. The fact that both repressor only models produced output levels similar to those for the induced system when Pc activity was set to zero is exactly the result expected physiologically. For a repressor only controlled gene, absence or inactivation of the repressor is sufficient to begin enzyme production provided that other necessary inputs are present. This occurs in the two repressor only models when the activity of state x_2 is set to zero.



Uninduced Response of GENE4A

Figure 5.1



Induced Response of GENE4B

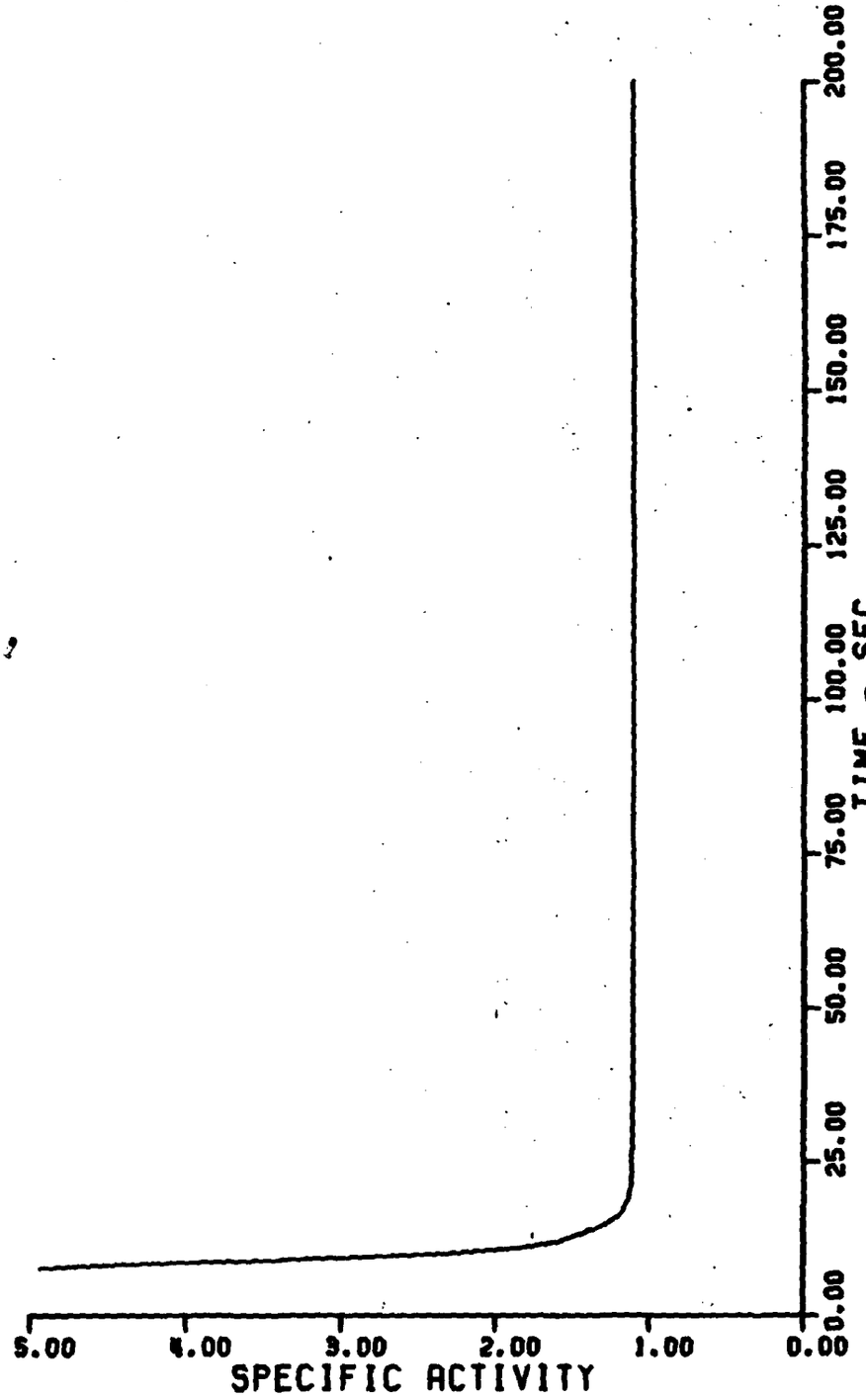
Figure 5.2

$$\dot{X}=\begin{bmatrix} -1 & 0 & -1.93 & 1 & 0 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ -.5 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix} X + \begin{bmatrix} 0 & .7 & 0 \\ 0 & .7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2.3 \\ -1.6 & 0 & 0 \end{bmatrix} U$$

$$Y = [37.4672 \quad 0 \quad 0 \quad 0 \quad 0] \; X$$

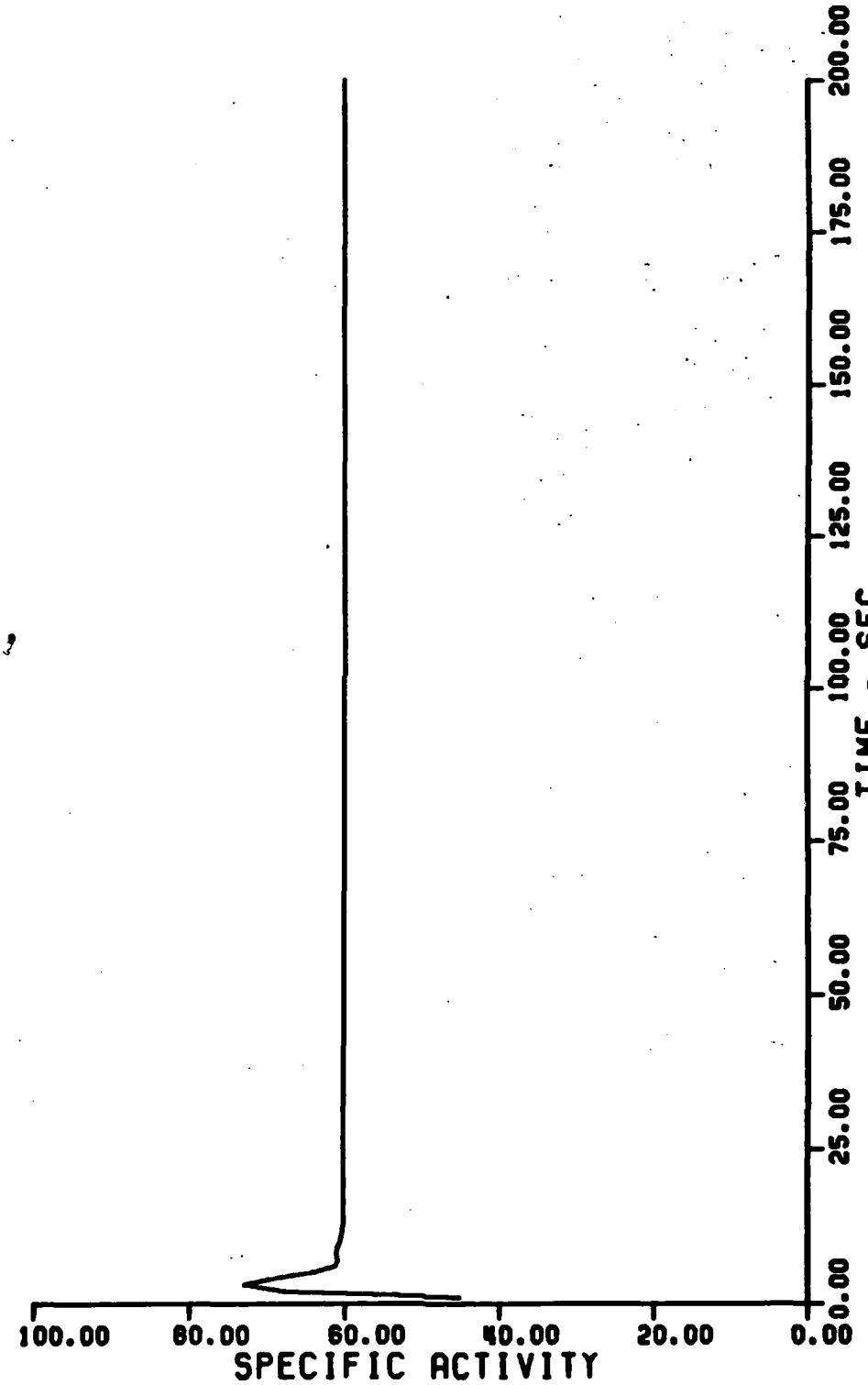
GENE8A/8B

Figure 5.3



Uninduced Response of GENEBA

Figure 5.4



Induced Response of GENE8B

Figure 5.5

Table 5.1

GENE4A/4B

B Matrix Parameters

Nominal Values: Y_U=29.97375 Y_I=44.96063

Coefficient	%Change	Y_U	$\%Y_U$	Y_I	$\%Y_I$
b_1 .5	-50	14.98683	50	29.97375	33.3
1.5	+50	44.96063	50	59.94751	33.3
b_2 .5	-50	37.46719	25	50.49996	12.3
1.5	+50	22.48032	25	37.46719	16.7
b_3 .5	-50	22.48032	25	37.46719	16.7
1.5	+50	37.46719	25	52.45407	16.7
b_4-.5	-50			37.46719	16.7
-1.5	+50			50.49996	12.3

Coefficeint	Y_U	e_U	Y_I	e_I
b_1 1.5			59.94751	.00275
1.51			60.24723	.06112
b_2 3.92	1.19898	.03959		
3.93	1.04912	.00241		
3.94	.89926	.01014		
b_4 3.92			59.79761	.04096
3.93			59.94747	.00276
3.94			60.09732	.00947

Table 5.2

GENE4A/4B
A Matrix Sensitivity Analysis

Nominal Values: Y_U=1.04908 Y_I=59.94751

Coefficient	%Change	Y_U	$\%Y_U$	Y_I	$\%Y_I$
a_2 -.9	-10	8.27217	68	60.24137	.5
-1.1	+10	0	100	59.64168	.5
a_3 .9	-10	0	100	59.32306	1
1.1	+10	3.89062	271	60.52389	1
a_5 -.9	-10	0	100	59.78807	.3
-1.1	+10	4.52410	331	60.08889	.2
a_6 .9	-10	2.49781	138	60.24137	.5
1.1	+10	0	100	59.64168	.5
a_8 .9	-10	4.88634	366	60.10362	.3
1.1	+10	0	100	59.80342	.2
a_{11} .9	-10	4.88634	366	63.24299	5.5
1.1	+10	0	100	54.14050	9.7

Table 5.3

GENE4A/4B

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.02800	1.60000	.65680
X2	2.45800	4.03000	3.08680
X3	2.45800	.10000	1.51480
X4	.98600	.20000	.67160
X5	2.45800	.10000	1.51480
Y	1.04908	59.94751	24.60844

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	16 sec.	6 sec.
5% Settling Time	13 sec.	4 sec.
10% Settling Time	11 sec.	2 sec.
Peak Overshoot	3718%	7.2%
Frequency	.17 Hz	.14 Hz
Time Constant	1.7 sec.	1.7 sec.

Table 5.4

GENE4A/4B Residual Analysis

Criterion	Expected	Computed	Residual	%Error
w/o Cat.Rep.	119.89502	74.93439	44.96063	37.50
0.4% ara	24.5	24.60844	.10844	.44
0.1 cAMP	23.97900	48.34948	24.37050	101.63
X1:X2 (I)	33.3:1	.4:1	32.9	98.80
0/2x cAMP	107.48283	74.93439	32.54844	30.28
X2=0	1.0	1.6666	.66667	66.67
10% S.T. (I)	300 sec.	2 sec.	298 sec.	
#-/0 values	0	0	0	

Table 5.5

GENE8A/8B

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.02985	1.60597	.66030
X2	1.49254	1.89851	1.65493
X3	1.49254	.29851	1.01493
X4	2.28507	1.49701	1.96985
X5	1.49254	.29851	1.01493
Y	1.11845	60.17121	24.73955

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	22 sec.	10 sec.
5% Settling Time	18 sec.	6 sec.
10% Settling Time	16 sec.	5 sec.
Peak Overshoot	4341%	21.3%
Frequency	---	---
Time Constant	2 sec.	1.9 sec.

Table 5.6

GENE8A/83 Residual Analysis

Criterion	Expected	Computed	Residual	%Error
w/o Cat.Rep.	120.34242	60.47208	59.87034	49.75
0.4% ara	24.5	24.73955	.23955	.98
0.1 cAMP	24.06848	31.73360	7.66512	31.85
X1:X2 (I)	33.3:1	.83:1	32.45	97.45
0/2x cAMP	66.64563	61.02869	5.61694	8.43
X2=0	1.0	74.93439	73.93439	7393.
10% S.T. (I)	300 sec.	16 sec.	284 sec.	
#-/0 values	0	0	0	

Chapter 6

Cost Analysis

6.1 Introduction

With state space models developed for both the activator-repressor and repressor only case, the hypothetical energy cost to the cell could be examined. Various energy expenditures are necessary for maintaining the DNA bindings and the concentrations of the controlling substances. The various states of the models should be related directly to the energy expended. Thus, it is proposed that a representation of the energy required for the operation of a gene can be calculated from the following equation.

$$J = X^T Q X + U^T R U \qquad (6.1)$$

where: J= Hypothetical cellular energy cost

Q= nxn matrix that assigns weights to the states of the system.

R= pxp matrix that assigns weights to the inputs of the system

If we assign equal weights to the states of the system and assume no cost for interactions among the states, then Q reduces to the identity matrix of order n. Also, if we assume no cost to the cell for the inputs, then R becomes a zero matrix, and the equation reduces to :

$$J = \Sigma x_i^2 \qquad (6.2)$$

Using this equation, the cost of regulation to the cell was calculated for GENEOA/OB, GENE4A/4B and GENE8A/8B. Table 6.1 presents the results of this cost analysis for the various system operations.

6.2 Cost analysis for the L-arabinose system

The results seem to indicate that, in these system situations, the repressor only models must expend more energy for similar system operation than the activator-repressor model. Initially, the analysis was done upon GENEOA/OB and GENE4A/4B only. The high costs associated with GENE4A/4B tended to be due to the high activity of x_2. It was then decided to redesign the repressor only model in an attempt to reduce the x_2 activity and hence lower the energy cost to the cell. When the activity of x_2 was reduced, other state values were necessarily raised in order to provide the same input/output system response. This implies that the higher cost of the repressor only models is not an artifact of model development.

For this particular gene-enzyme complex, the more complicated control mechanism may be the most energy efficient for the cell. An increase in the number of controlling sites along the DNA may reduce the overall DNA activity necessary for gene operation compared to a structure in which the control activity is restricted to a smaller number of controlling sites on the DNA.

The "Demand Theory of Gene Regulation" suggests that in an arabinose rich environment, activator control would be selected for while in an arabinose poor environment, repressor control would be selected for. It further suggests that the dual control of the system has been selected for because E. coli normally inhabits both environments about equally during its life time. Thus, this gene-enzyme complex could be expected to be utilized over about one-half of the organism's life span and then not used for the other half. Since the system may be off about 50% of the time, it would be desirable to have the repression of the system be inexpensive. This appears to be the case for the dual control model of the L-arabinose system, and not the case for the repressor only model. The dual of this concept should also be true. That is, the operational cost in the induced system should be less than in an activator only controlled system. Analysis of this problem must await development of an activator only model for the L-arabinose system.

Under conditions that reduce the effective concentration of the controlling protein, the dual control model also appears to cost less than the repressor only model. Mutations that reduce the effectiveness of the protein are far more common than mutations that increase the effectiveness of it. If the organism were found only in an arabinose rich environment, these mutations in a repressor only system would not be selected against because enzyme production would continue unabated and the arabinose utilized. The organism, however, is not always in an arabinose rich environment and thus continued enzyme production is

wasteful and would be selected against. Thus, the cheaper cost of the dual control model under these circumstances would appear to be advantageous for the organism.

Thus, it appears that the above cost analysis supports the theory that this particular dual control system has been selected for based upon the demands of the environment for the unique operation of this gene-enzyme complex.

Following development of the models, it was suggested that the power of comparisons such as those done here is greatest if the systems being compared differ by only one parameter [35]. An attempt to remodel the repressor only system using such an approach yielded the model, GENE9A/9B, shown in Figure 6.1.

The A matrix of this model was restructured in an attempt to make this model as close to GENE0A/0B as possible. As shown in Figure 6.1, corresponding coefficients of the two models are identical with the exception of the coefficient representing repression by ara0 which is -1.98 in GENE9A/9B and -1.0 in GENE0A/0B. Note that the state feedback of x_5 upon itself which was used to control the concentration of the repressor in GENE4A/4B and GENE8A/8B has been replaced with two terms, one representing the loss of repressor when bound to x_3 and one representing protein degradation which is analogous to the conversion of P1 to P2 in GENE0A/0B.

The system responses of GENE9A/9B are presented in Tables 6.2 and 6.3 and in Figures 6.2 and 6.3. The responses of this model are consistent with the models developed earlier. In addition, this model

shows improved responses when cAMP input is varied.

A detailed cost analysis comparing GENE0A/0B and GENE9A/9B is presented in Table 6.4. The data in Table 6.4 reveals that following mutations that would yield no repressor or an inoperable repressor (X2=0 or X5=0), the complex form of control saves the cell from unnecessary energy expenditure. The costs for the induced system are similar although distributed differently between the two models. The major difference between the two models is in the uninduced cost. As revealed in Table 6.4, the repressor only model has a substantially higher cost than the activator/repressor model with the higher cost distributed fairly equally over all states except x_1. This higher cost in the uninduced system was found in all repressor only models. As mentioned earlier, this indicates that for the L-arabinose system, complex control may be more cost efficient than simple control.

A plot (Figure 6.4) was made of the energy cost to the organism, as computed from GENE0A/0B and GENE9A/9B, versus the percentage of time that the L-arabinose system in induced. Thus, the cost of dual control was compared to the cost of repressor only control over the full range of induction possibilities.

As shown in Figure 6.4, the data indicates that the L-arabinose system would have to be induced greater than 85 percent of the life span of the organism in order for repressor only control to be less energy expensive to the organism. Thus, for E. coli, an organism that can expect to use this operon fifty percent of the time [1], the complex dual control appears to be more cost effective than a simple

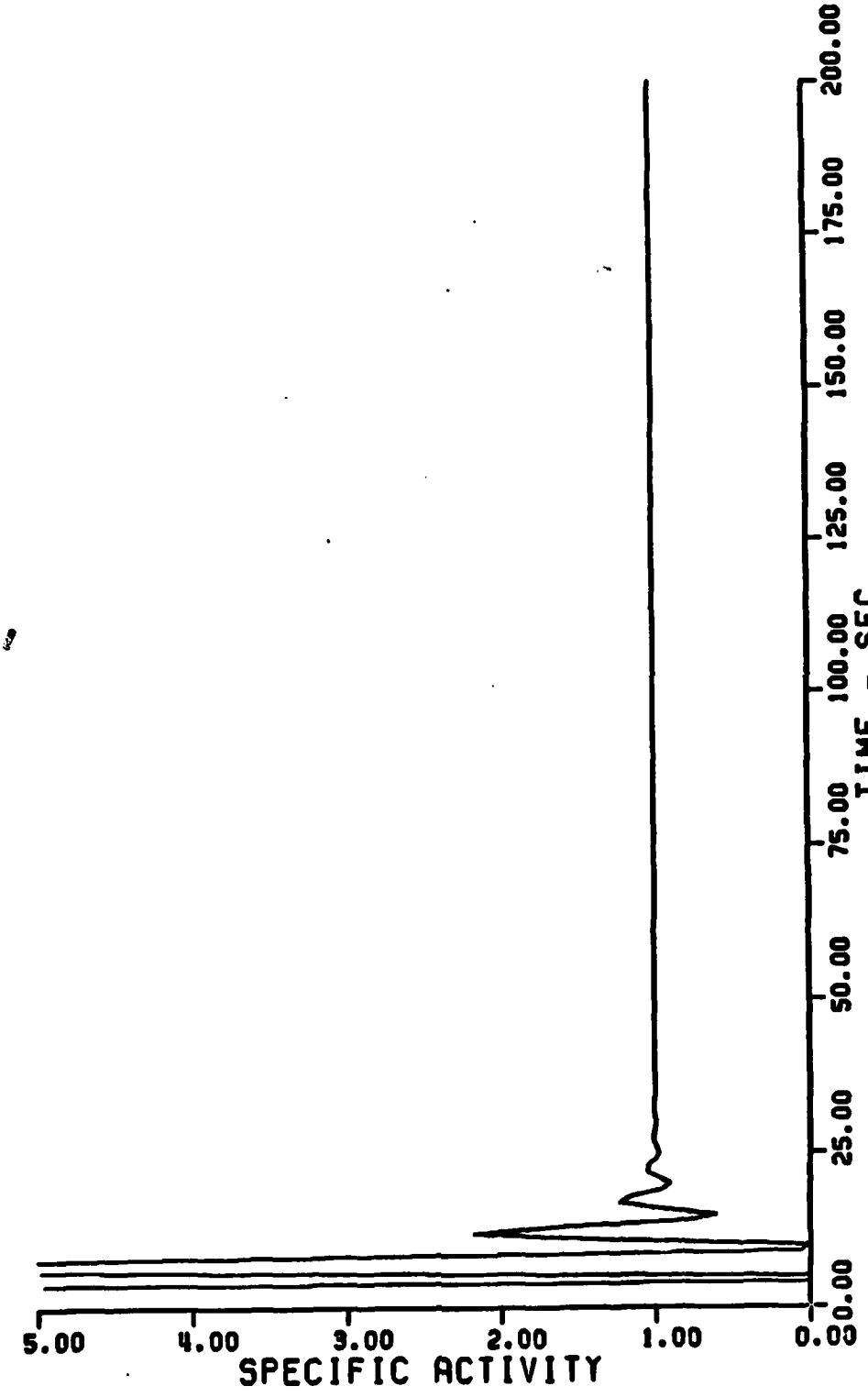
repressor only control.

$$\dot{X}=\begin{bmatrix} -1 & 0 & -1.98 & 1 & 0 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ -.5 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & -.001 \end{bmatrix} X + \begin{bmatrix} 0 & .7 & 0 \\ 0 & .7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1.86 \\ -1.6 & 0 & 0 \end{bmatrix} U$$

$$y = [37.4672 \quad 0 \quad 0 \quad 0 \quad 0] \quad X$$

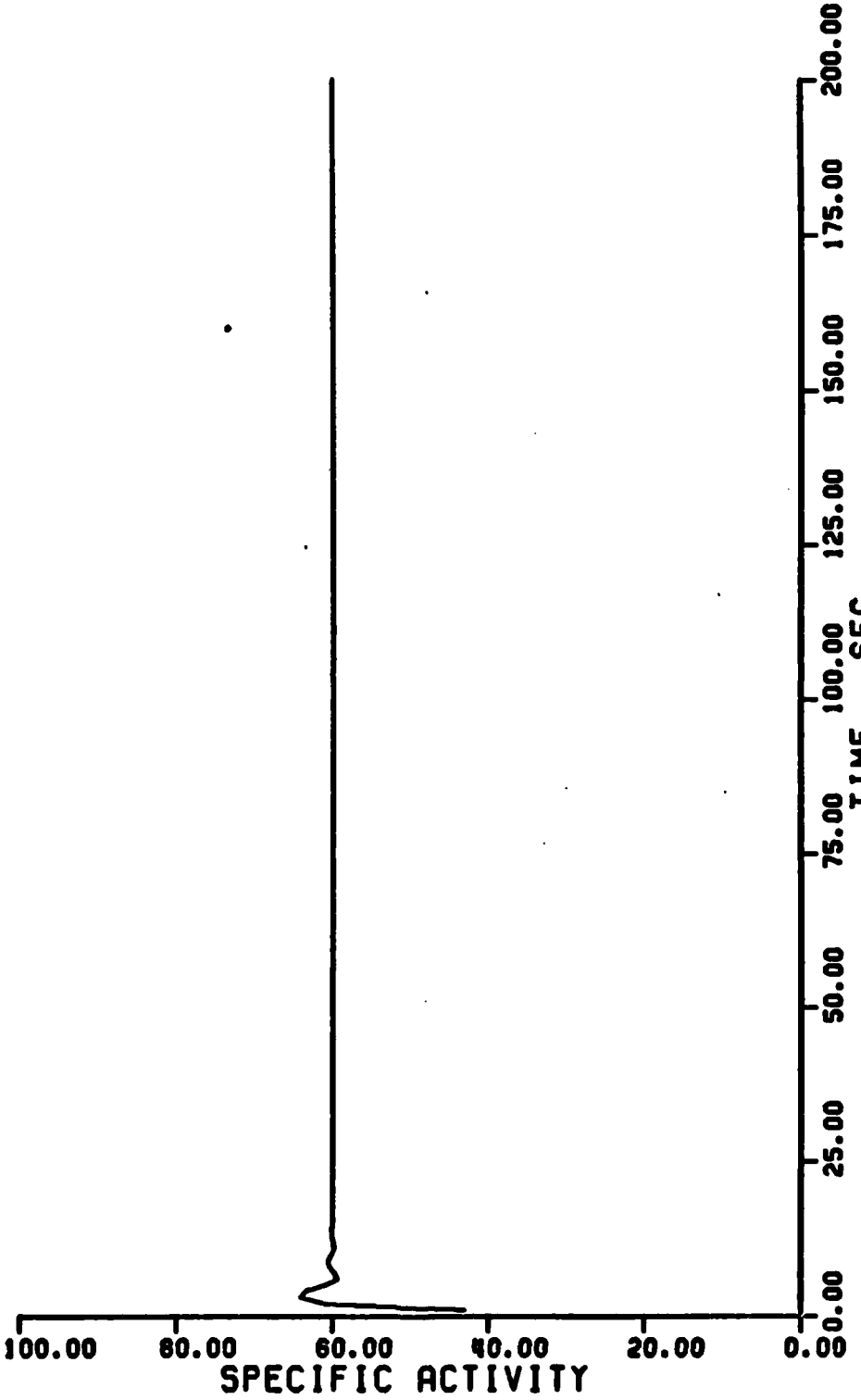
GENE9A/9B

Figure 6.1



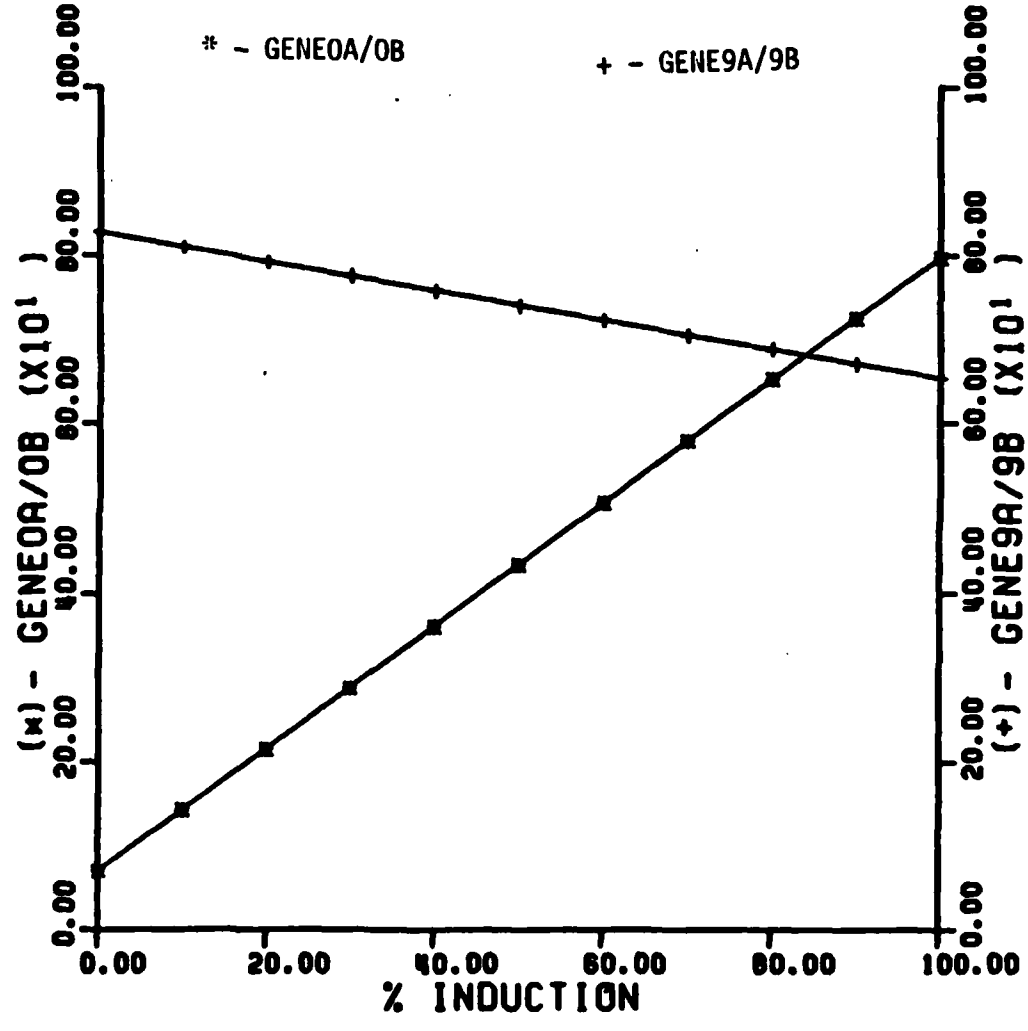
Uninduced Response of GENE9A

Figure 6.2



Induced Response of GENE9B

Figure 6.3



Energy Cost vs Percent Induction

GENEOA/OB (Eight State Model) & GENE9A/9B (Five State Model)

Figure 6.4

Table 6.1

Cost Analysis

Criterion	GENE0A/0B	GENE4A/4B	GENE8A/8B
Uninduced	.6979	19.098	11.905
Induced	7.9677	18.861	8.603
X2 = 0	.1151	2.806	5.690
X7=X8=0 (X5)	.6517	19.588	9.690

State Values:

	Uninduced			Induced		
State	GENE0A/0B	GENE4A/4B	GENE8A/8B	GENE0A/0B	GENE4A/4B	GENE8A/8B
X1	.02669	.02800	.02985	1.59421	1.60000	1.60597
X2	.50077	2.45800	1.49254	.07694	4.03000	1.89851
X3	.50096	2.45800	1.49254	0	.10000	.29851
X4	0	.98600	2.28507	1.63394	.20000	1.49701
X5	.02712	2.45800	1.49254	1.63869	.10000	.29851
X6	.34659	-------	-------	0	------	------
X7	.26215	-------	-------	.04654	------	------
X8	.07663	-------	-------	.25099	-------	------

Table 6.2

GENE9A/9B

Nominal Values:

	Uninduced System	Induced (1%)	Induced (.4%)
X1	.02673	1.60167	.65670
X2	1.27395	1.67962	1.43622
X3	1.27268	.07954	.79543
X4	1.84664	1.05916	1.53165
X5	1.27268	.07954	.79543
Y	1.00138	60.01011	24.60487

e_U=.00000 e_I=.00010

Transient Analysis:

Characteristic	Uninduced System	Induced (1%)
1% Settling Time	26 sec.	7 sec.
5% Settling Time	21 sec.	5 sec.
10% Settling Time	21 sec.	1 sec.
Peak Overshoot	3751%	6.9%
Frequency	.2 Hz	.2 Hz
Time Constant	1.2 sec.	1.7 sec.

Table 6.3

GENE9A/9B Residual Analysis

Criterion	Expected	Computed	Residual	%Error
w/o Cat.Rep.	120.02022	60.32501	59.69521	49.74
0.4% ara	24.5	24.60487	.10487	.43
0.1 cAMP	24.0040	32.73021	8.72617	36.35
X1:X2 (I)	33.3:1	.95:1	32.35	97.14
0/2x cAMP	70.20033	60.73768	9.46265	13.48
X2=0	1.0	63.94400	62.94400	6294
10% S.T. (I)	300 sec.	1 sec.	299 sec.	
#-/0 values	0	0	0	

Table 6.4

Cost Analysis

Criterion	GENEOA/OB	GENE9A/9B
Uninduced	.6979	8.27313
Induced	7.9677	6.52096
X2 = 0	.1151	3.92609
X7=X8=0 (X5)	.6517	6.83880

State Values:

	Uninduced		Induced (1%)	
State	GENEOA/OB	GENE9A/9B	GENEOA/OB	GENE9A/9B
X1	.02669	.02673	1.59421	1.60167
X2	.50077	1.27395	.07694	1.67962
X3	.50096	1.27268	0	.07954
X4	0	1.84664	1.63394	1.05916
X5	.02712	1.27268	1.63869	.07954
X6	.34659	---	0	---
X7	.26215	---	.04654	---
X8	.07663	---	.25099	---

Chapter 7

Discussion

7.1 Conclusions

The L-arabinose Gene-Enzyme complex was modeled as a linear time invariant system. The model which best fit the data, GENEOA/OB, consisted of eight states representing the activity of the controlling sites on the DNA and the concentrations of the controlling proteins (repressor and activator).

The input/output behavior of the model was compared to data from the literature. The output values for the model corresponded well to those values for the biological system, reported in the literature for the uninduced system, the system induced with 1% arabinose and the system induced with 0.4% arabinose. All of these output values were well within the range of values reported in the literature.

The values of the states of the system also corresponded well with the reported activities of the controlling sites on the DNA. These values included the PBAD:Pc ratio (x1:x2), the P2/araO:P2/araI ratio (x4:x5) and the P1/araO:P2araO ratio (x3:x4).

The model responded to the removal of catabolite repression with an increase in enzyme production, similar to the response of the biological system although not to the extent reported in the literature. Also, the model responded to varying inputs of cAMP in a

manner similar to the biological system. That is, a 90% decrease in cAMP input caused a 1.5% decrease in the output of the model, and an increase in cAMP input by a factor of two caused a 3.3% increase in the output of the model. Once again, these changes in the output were not as large as those reported in the literature.

The output of the model reached a steady state value well within the reported response time of the biological system (5 minutes), and state dependencies were displayed that were predictable from the experimental data as shown in Figures 4.3 - 4.16.

The model was found to be asymptotically stable and fully controllable through the inputs as predicted by the data on the biological system.

A comparison was made between the eight state dual control model and a five state hypothetical repressor only model of the L-arabinose system. The comparison was based upon the activities of the states of the system which were related to the energy cost to the cell. When the two models are fitted to biological data, results of the comparison indicate that the more complex eight state model has a lower energy cost than the five state model, requiring about one half as much energy. This implies that for the L-arabinose Gene-Enzyme complex, the more complex dual method of control may be more energy efficient than a simple repressor only method of control.

This result supports the "Demand Theory of Gene Regulation" by suggesting that selection favored the method of control that was most energy efficient for an operon that must be active about 50% of the time.

7.2 Discussion of the Models

The approach of this research in many respects was unique. The formulation of the equations describing the system's behavior was based where possible upon relationships between the states of the system as suggested in the literature, and not upon chemical kinetic equations. Contributing to this choice, was the lack of available quantitive data on the DNA kinetics of the L-arabinose system, and of data on genetic control systems in general. In addition, these equations assumed a totally linear relationship of variables. The philosophy of the analysis was that a first approximation based upon linearity could later be improved by the addition of non linearities within the format of the model. In addition, a larger number of standard analysis techniques are available for a linear model than for a nonlinear one. For a first approximation, the model adequately mimiced the actual system for normal input/output responses, physiological parameters, such as activity ratios for various DNA controlling sites, and mutational behavior, such as when there is no activity of the controlling gene. Results in those areas fell within the published experimental data. The x_1:x_2 activity ratio in the induced system was approximately 2:1 and the x_4:x_5 ratio

was nearly 1:1. When the controlling gene was inactive, the induced output remained near the value for the uninduced system.

As discussed before, one area where the models encountered difficulty was in the responses of the system to removal of catabolite repression and varying amounts of cAMP. This may be due in part to the basic model structure or to the linearity of the model. cAMP is involved in many cellular activities, as well as being a part of the regulatory process for many genetic systems [36,37]. Thus, because cAMP interacts in many different areas, the modeling of its binding in this specific system, without accounting for other interactions, may present more difficulty than that encountered for other substances.

Certain aspects of the system's behavior were not accounted for by using the models. A temporary derepression of the araC promoter immediately following induction has been reported in the literature [29,18]. The responses of the models do not demonstrate this behavior, suggesting that additional complexities must be built into the model to account for this behavior.

7.3 Future Considerations

As mentioned, all models encountered difficulty when catabolite repression was removed or when cAMP input was varied. Possible changes in the model were proposed that may improve the responses of the model. These changes included the addition of another state to model cAMP concentration. Incorporation of these or other changes are vital to further use of the model.

Additional changes to the model are necessary to provide the temporary derepression of araC that has been observed immediately following induction [29]. It has been suggested that this derepression is dependant upon L-arabinose concentration, such that small amounts of arabinose cause P1 to dissociate from araO allowing araC to be more active until the concentration of arabinose is great enough to form sufficient P2 to bind both araI and araO, and once again repress araC [29]. Simulating this effect most likely will involve changes to the equations for Pc, araO, and araI activity as well as for those involving P1 and P2 concentrations.

Changes to the input and output formulations may also be necessary. The transport of L-arabinose into the cell could be considered as well as the internal production and utilization of cAMP and RNAP. Also, the relationship between PBAD activity and enzyme production should be reevaluated.

GLOSSARY

araA, araB, araC - Structual genes of the L-arabinose system

araC - Controlling gene for the L-arabinose system

araI - Initiator site/binding site for the activator

araO - operator site/Binding site for the repressor

cAMP - cyclic adenosine monophosphate

catabolite repression - The decrease in free cAMP attributed to the increase in metabolic products.

CRP - cAMP receptor protein

GENE / - The various state space models of the system.

Induced system - System with L-arabinose input

L-arabinose - A five carbon sugar

Least square error - (Yexpected-Ycomputed) squared

LTI - Linear time invariant

Percent error - (residual/expected value) x 100

P1 - Repressor protein

P2 - Activator protein

PBAD - Promoter for araBAD/site of RNAP attachment

Pc - Promoter for araC/site of RNAP attachment

Residual - expected value minus computed value

RNAP - RNA polymerase/binds to DNA to begin transcription

Uninduced - System with no L-arabinose input.

Small letters represent individual states or scalers.

Capital letters represent vectors or matrices.

LIST OF REFERENCES

1. Savageau, M. A. (1983) "Regulation of Differentiated Cell-Specific Functions" Proc. Natl. Acad. Sci. USA 80,1411-1415.

2. Savageau, M. A. (1982) "Models of Gene Function" in Foundations of Biochemical Engineering : Kinetics and Thermodynamics in Biological Systems, eds. Papoutsokis, T., Stephanopoulos, G. W. and Blanch, H. W. (American Chemical Society, Washington, D.C.), pp. 1-24.

3. Kwakernaak, H. and Sivan, R. Linear Optimal Control Systems. Wiley-Interscience, New York, New York. 1972.

4. Englesberg, E. and Wilcox, G. (1974) "The L-arabinose Gene-Enzyme Complex" Annu. Rev. Genetics 8, 219-242.

5. Lee, N. (1978) "Molecular Aspects of ara Regulation" in The Operon, eds. Miller, J. H. and Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), pp. 389-409.

6. Lewin, B., "Genes." Wiley Interscience, N.Y., N.Y. 1983.

7. Sheppard, D. and Englesberg, E. (1967) "Further Evidence for Positive Control of the L-arabinose System by Gene araC" J. Mol. Biol. 25,443-454.

8. Iyengar, S. S. "Computer Modeling of Complex Biological Systems." CRC Press, Boca Raton, Florida. 1984.

9. Economos, Angelos C. (1980) "Dynamic Studies and Modelling of Physiological Systems" Cybernetica 23,229-257.

10. Schleif, R. (1972) "Structure Deletion Map of the E. coli L-arabinose Operon" Proc. Natl. Acad. Sci. USA 69, 3479-3484.

11. Schleif, R. and Lis, J. (1975) " The Regulatory Region of the L-arabinose Operon, A Physical, Genetic and Physiological Study" J. Mol. Bio. 95, 417-431.

12. Watson, James D. Molecular Biology of the Gene. W. A. Benjamin, Menlo Park, California 1976.

13. Englesberg, E., Irr, J., Power, J., and Lee, N. (1965) "Positive Control of Enzyme Synthesis by Gene C in the L-arabinose System" J. Bacteriol. 90, 946-956.

14. Raibaud, O. and Schwartz, M. (1984) "Positive Control of Transcription Initiation in Bacteria" Ann. Rev. Genetics 18, 173-206.

15. Lee, N. L., Gielow, W. O. and Wallace, R. G. (1981) "Mechanism of araC Autoregulation and the Domains of Two Overlapping Promoters, Pc and PBAD, in the L-arabinose Regulatory Region of Escherichia coli" Proc. Natl. Acad. Sci. USA 78, 752-756.

16. Dunn, T. M., Hahn, S., Ogden, S. and Schleif, R. F. (1984) "An Operator at -280 Base Pairs That is Required for Repression of araBAD Operon Promoter: Addition of DNA Helical Turns Between the Operator and Promoter Cyclically Hinders Repression" J. Mol. Biol. 174, 611-628.

17. Hahn, S. and Schleif, R. (1983) "In Vivo Regulation of the Escherichia coli araC Promoter" J. Bacteriol. 155, 593-600.

18. Ogden, S., Haggerty, D. and Schleif, R. (1980) "The Escherichia coli L-arabinose Operon: Binding Sites of the Regulatory Proteins and a Mechanism of Positive and Negative Regulation" Proc. Natl. Acad. Sci. USA 77, 3346-3350.

19. Hendrickson, W. and Schleif, R. (1984) "Regulation of the Escherichia coli L-arabinose Operon Studied by Gel Electrophoresis DNA Binding Assay" J. Mol. Biol. 174, 611-628.

20. Savageau, M. A. Biochemical Systems Analysis Addison-Wesley, Reading, Massachusetts. 1976.

21. Tapaswi, P. K. and Bhattacharya, P. (1981) "An Extended Mathematical Model of Transcription and Translation During Embryogenesis" Cybernetica 24, 61-84.

22. Tchuraev, R. N. and Ratner, V. A. (1983) "A Continous Approach With Threshold Characteristics for Simulation of Gene Expression" in Molecular Genetic Information Systems Modelling and Simulation, ed. Bellmann, Klaus (Akademie-Verhag, Berlin) pp. 64-80.

23. Alberghina, L. and Martegani, E. (1979) "Ribosome and Protein Synthesis, Chromosome Replication, and Cell Division in Escherichia coli" Cybernetica 22, 57-81.

24. Kailath, Thomas Linear Systems Prentice-Hall Englewood Cliffs, N. J. 1980.

25. Haggerty, D., Oeschger, M. and Schleif, R. (1978) "In Vivo Titration of araC Protein" J. Bacteriol. 135, 775-781.

26. Neumann, H. and Kreischer, U. (1983) "Mathematico-Cybernetical Model of Prokaryotic Transcription Systems" in Molecular Genetic Information Systems Modelling and Simulation ed. Bellmann, Klaus (Akademie-Verhag, Berlin) pp. 133-165.

27. Bellmann, K., Born, J., Lindigkeit, R. and Schultz, R. (1983) "Modelling and Simulation in Analysis of Eukaryotic Transcription Control" in Molecular Genetic Information Systems Modelling and Simulation ed. Bellmann, Klaus (Akademie-Verhag, Berlin) pp. 234-266.

28. Silverman, Leonard M. (1971) "Realization of Linear Dynamical Systems" IEEE Trans Automatic Control 16, 554-567.

29. Casadaban, M. (1976) "Regulation of the Regulatory Gene for the L-arabinose Pathway, araC" J. Mol. Biol. 104, 557-566.

30. Englesberg, E.,Sheppard, D., Squires, C. and Meronk, F. (1969) "An Analysis of "Revertants" of a Deletion Mutant in the C Gene of the L-arabinose Gene Complex in Escherichia coli Br: Isolation of Initiator Constitutive Mutants" J. Mol. Biol. 43, 281-298.

31. Katz, L. and Englesberg, E. (1971) "Hyperinducibility as a Result of Mutation in Structural Genes and Self-Catabolite Repression in the ara Operon" J. Bacteriol. 107, 34-52.

32. Lis, J. and Schleif, R. (1973) "Different Cyclic AMP Requirements for Induction of the Arabinose and Lactose Operons of Escherichia coli" J. Mol. Biol. 79, 149-162.

33. Heffernan, L. and Wilcox, G. (1976) "Mutations Affecting Catabolite Repression of the Larabinose Regulon in Escherichia coli B/r" J. Bacteriol. 126, 1132-1135.

34. Brockett, R.W. and Byrnes, C. I. (1981) "Multivariable Nyquist Criteria, Root Loci, and Pole Placement: A Geometric Viewpoint" IEEE Trans. Auto. Control 26, 271-283.

35. Savageau, M. A. (Personal Communication)

36. Gottesman, S. (1984) "Bacterial Regulation: Global Regulatory Networks" Ann. Rev. Genetics 18, 415-441.

37. deCrombrugghe, B., Busby, S. and Buc, H. (1984) "Cyclic AMP Receptor Protein: Role in Transcription Activation" Science 224, 831-837.

END

DTIC

5-86